地理信息系统理论与应用丛书

吴立新　史文中　编著

地理信息系统原理与算法


科学出版社
www.sciencep.com

(P-1420.0101)
责任编辑：彭胜潮　姚岁寒



吴立新博士（1966－），江西宜春人，中国矿业大学（北京校区）测绘与土地科学系教授、博士生导师，3S与沉陷工程研究所所长。1997年在中国矿业大学大地测量学与测量工程专业获博士学位；中国煤炭青年科技奖、全国百篇优秀博士学位论文、中国高校青年教师奖获得者，《地理与地理信息科学》副主编。目前研究领域包括：3DGIS理论与方法、数字矿山、数据挖掘、遥感-岩石力学、开采沉陷与防护、可持续发展等。已发表学术论文170余篇，出版专著3部、教材1部、译著1部。

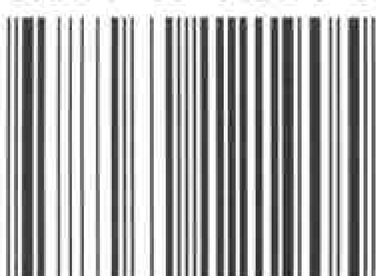
Dr Wu Lixin, Director, Institute of RS/GPS/GIS and Subsidence Engineering, and Professor, Department of Geomatics and Land Science, China University of Mining & Technology (Beijing). Current research interests include theory and algorithms in GIS, 3D Geosciences modeling, spatial data mining, digital mine, remote sensing rock mechanics, earthquake remote sensing, INSAR monitoring for subsidence, mining subsidence control and precaution, and sustainable development.
E-mail: awulixin@263.net; wwulixin@sina.com; Tel: +086 10 62331012



史文中博士（1963－），江苏无锡人，香港理工大学土地测量及地理资讯系副教授，地球咨询科技研究中心主任。1994年在德国Osnabrueck大学获博士学位，研究方向为GIS与RS；曾任香港GIS学会主席，现任ISPRS IV-1工作组组长，应邀为IJRS、ISPRS J. PRS、IJGIS、PE&RS、Transactions in GIS、Geoinformatica等国际刊物审稿或任客座主编。目前研究领域包括：空间数据质量、3DGIS数据模型、虚拟现实、GIS与RS集成、GIS设计与开发、RS影像特征提取等。已发表学术论文200余篇，出版专著4部。

Dr Shi Wenzhong, Director, Advanced Research Centre for Spatial Information Technology, and Associate Professor, Department of Land Surveying and Geo-Informatics, the Hong Kong Polytechnic University. Current research interests include GIS, remote sensing and virtual reality, spatial data quality, 3D and dynamic data model in GIS, design and development of GIS, integration of GIS and remote sensing, and feature extraction from remotely sensed images.
E-mail: lswzshi@polyu.edu.hk ; Tel: + 852 - 2766 5975

ISBN 7-03-012190-2



9 787030 121905 >

ISBN 7-03-012190-2
定 价：60.00 元

地理信息系统理论与应用丛书

地理信息系统原理与算法

吴立新　史文中　编著

教育部第三届“高校青年教师奖”支持计划
香港政府研究资助局项目(3_ZB40)
中国矿业大学研究生教材出版基金
香港理工大学基金(1.34.9709)
联合资助

科学出版社
北京

内 容 简 介

本书系统总结、详细整理了国内外GIS理论与核心算法的成果,并进行归纳分类、比较分析和提炼演绎。全书以GIS原理和GIS算法为纲,形成"GIS原理"和"GIS算法"上、下两篇。上篇基于学科融合与交叉提出了广义地理系统和广义地理目标的概念,进而系统讲解了2D GIS空间数据模型、数据结构、空间关系、地学信息分类与空间编码、地图投影与图形变换等基本原理,并介绍了近年GIS理论的最新进展,如3D GIS空间数据模型、时态GIS及时态关系和地理元数据。下篇基于空间数据处理、空间图形处理、空间度量与分析、可视化与数据挖掘共4大方面,系统阐述和演绎了空间数据压缩、空间数据内插、空间数据转换、空间数据误差分析算法,多边形自动生成与裁剪、TIN的构建、Voronoi图的构建和空间变换算法,空间度量、数字地形分析、空间统计和空间分析算法,以及GIS可视化操作和空间数据挖掘与知识发现等核心算法。

本书可作为GIS相关专业的硕士生、博士生教材,也可作为GIS高级研究人员、GIS系统研发人员和GIS高级程序员的参考用书和技术指南。

图书在版编目(CIP)数据

地理信息系统原理与算法/吴立新,史文中编著.—北京:科学出版社,2003

(地理信息系统理论与应用丛书)

ISBN 7-03-012190-2

Ⅰ.地… Ⅱ.①吴… ②史… Ⅲ.地理信息系统 Ⅳ.P208

中国版本图书馆CIP数据核字(2003)第085815号

责任编辑:彭胜潮 姚岁寒/责任校对:宋玲玲

责任印制:钱玉芬/封面设计:王 浩

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

http://www.sciencep.com

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2003年10月第 一 版 开本:787×1092 1/16

2003年10月第一次印刷 印张:29 1/2

印数:1—3 500 字数:654 000

定价:60.00元

(如有印装质量问题,我社负责调换〈新欣〉)

序　言

一部有关 GIS 算法的著作的出版，是我期盼已久的事情。

自 20 世纪 90 年代以来，我国 GIS 理论、方法、技术与应用研究蓬勃发展，GIS 教育更是蒸蒸日上，尤其是 GIS 方向的硕士、博士研究生的培养规模日益扩大。GIS 原理与算法已成为 GIS 高级研究的焦点和 GIS 基础软件开发的关键。长期以来，国内一直缺乏一部系统阐述 GIS 核心算法的著作或研究生教材；国外也只是在 1997 年才由荷兰学者出版了一部名为《GIS 算法基础》(Algorithmic Foundations of GIS) 的书（16 开，287 页)。该书重点阐述了 GIS 中的 Voronoi 方法、DEM 和 TIN 算法、空间数据综合算法、内存扩展算法等有限内容。

吴立新、史文中两位教授系统总结、详细整理了国内外 GIS 理论与核心算法的成果，并进行归纳分类、分析比较和提炼演绎，同时结合中国内地和香港地区 GIS 研究生教育的经验和实际，并融合双方近年在 GIS 理论与算法方面的研究成果，完成了这部著作。纵观全书，以 GIS 原理和 GIS 算法为纲形成上、下两篇，结构清晰，体系完整。

上篇“GIS 原理”部分不仅系统讲解了 2D GIS 空间数据模型、数据结构、空间关系、地学分类与空间编码、地图投影与图形变换等 GIS 原理的基本问题，而且纳入了近年 GIS 理论的最新发展，如 3D GIS 空间数据模型、时态 GIS 及时态关系和地理元数据。尤其，从地理认知理论出发，扩充了地理系统的概念和 GIS 的学科范畴，基于学科融合与交叉把涉及地球内部、表面及外部的地理系统称为广义地理系统，进而提出了广义地理目标的新概念，并以此形成本书的基础与前提。

下篇“GIS 算法”部分从空间数据处理、空间图形处理、空间度量与分析、可视化与数据挖掘共 4 个板块出发，系统阐述和演绎了空间数据压缩、空间数据内插、空间数据转换、空间数据误差分析算法，多边形自动生成与裁剪、TIN 的构建、Voronoi 图的构建和空间变换算法，空间度量、数字地形分析、空间统计和空间分析算法，以及 GIS 可视化操纵和空间数据挖掘与知识发现算法。

可以说，该书的成果是国内外 GIS 理论与算法研究成果的集大成，是对我国 GIS 学术界和教育界的一大贡献。该书不仅可作为我国 GIS 研究生教育的公共教材，也可作为我国 GIS 高级开发人员的技术指南。相信，该书的出版，必将促进我国 GIS 研究、教育与技术的发展。同时，希望作者继续关

注本领域国内外研究动态和最新成果，随时对新理论、新算法进行总结、归纳和提炼，以便在本书后续再版中不断纳入进去，使本书的体系和内容不断丰富和完善。

中国科学院院士
中国工程院院士
李德仁

2003年6月27日于武汉

前　言

英国 Edinburgh 大学地理学教授 Terry Coppock 自 20 世纪 50 年代后期开始，就使用早期的计算机进行与土地利用变化及国土资源相关信息的收集、管理和操作工作，并于 20 世纪 80 年代中叶创刊了 *International Journal of Geographical Information Systems*（简称 *IJGIS*，现改名为 *International Journal of Geographical Information Science*）这一最具国际影响的 GIS 刊物，并担任首任编辑。前加拿大地理师协会主席、皇家地理学会 Murchison 奖获得者、测量学家 Roger Tomlinson，则在 20 世纪 60 年代认识到数字计算机可以用于加拿大土地调查署的海量地图信息的分析工作，并通过应用实践证明计算机是地理信息处理与分析的最佳选择，率先提出了 Geographical Information System (GIS) 的术语，进而成为 GIS 之父。美国加州大学地理系 Michael F. Goodchild 教授自 70 年代以来为推动国际 GIS 理论与技术发展作出了积极贡献；我国的陈述彭院士、李德仁院士等老一代科学家均为推动和发展中国的 GIS 理论、技术与产业做出了卓越贡献。地理信息技术经过 40 年的发展，已经形成国际 GIS 理论研究、技术开发、社会应用与产业化四项并举的欣欣向荣局面。

我国自 20 世纪 80 年代开始大力赶超以来，目前已基本达到与国际先进水平齐平，在某些方面甚至还有所超越。在理论研究方面，国内许多高校的大地测量学与测量工程（081601）、摄影测量与遥感（081602）、地图制图学与地理信息工程专业（081603）、矿山空间信息学与沉陷工程（081620）、自然地理学（070501）、人文地理学（070502）、地图学与地理信息系统（070503）、地球探测与信息技术（081802）、城市规划与设计（081303）、土地资源管理（120405）等二级学科均已经有了相应的博士点或硕士点，并进行与 GIS 相关的研究生培养和理论探索；中国科学院系统的地理科学与资源研究所、遥感应用研究所以及中国测绘科学研究院、国家基础地理信息中心等单位也在 GIS 理论研究方面取得显著成绩。技术开发方面，除中国高校和科学院系统之外，许多公司（包括外资公司、合资公司和独资公司等），均在 GIS 的基础开发与应用开发方面取得突出成绩。社会应用方面，GIS 的应用领域已经深入社会经济的各个层面、各个领域，如城市、乡村、气象、农业、林业、矿山、地质、水文、人口、保险、公安、交通、海关等；从地理专业应用，到行业应用、综合应用和公众应用，几乎“无所不包，无所不在，无所不能”。产业化方面，从 90 年代初期开始，我国 GIS 的产业化初见端倪；1994 年中国 GIS 协会（CAGIS）的成立和 1997 年中国科学技术协会第 25 次青年科学论坛（以“地理信息科学与产业化”为主题）的召开，有效地规范、引导和促进了我国 GIS 产业化发展。目前中国内地 GIS 的产业收入（包括软件销售、数据服务、技术开发、技术服务）已超过 30 亿元。

国际上自 GIS 经典之作 *Principles of GIS for Land Resources Assessment*（P. A. Burrouch, 1986）出版以来，GIS 方面的教材、著作、论文集等层出不穷。代表作有：*Three Dimensional Applications in GISs*（J. Raper, 1989），*GIS and Cartographic Mo-*

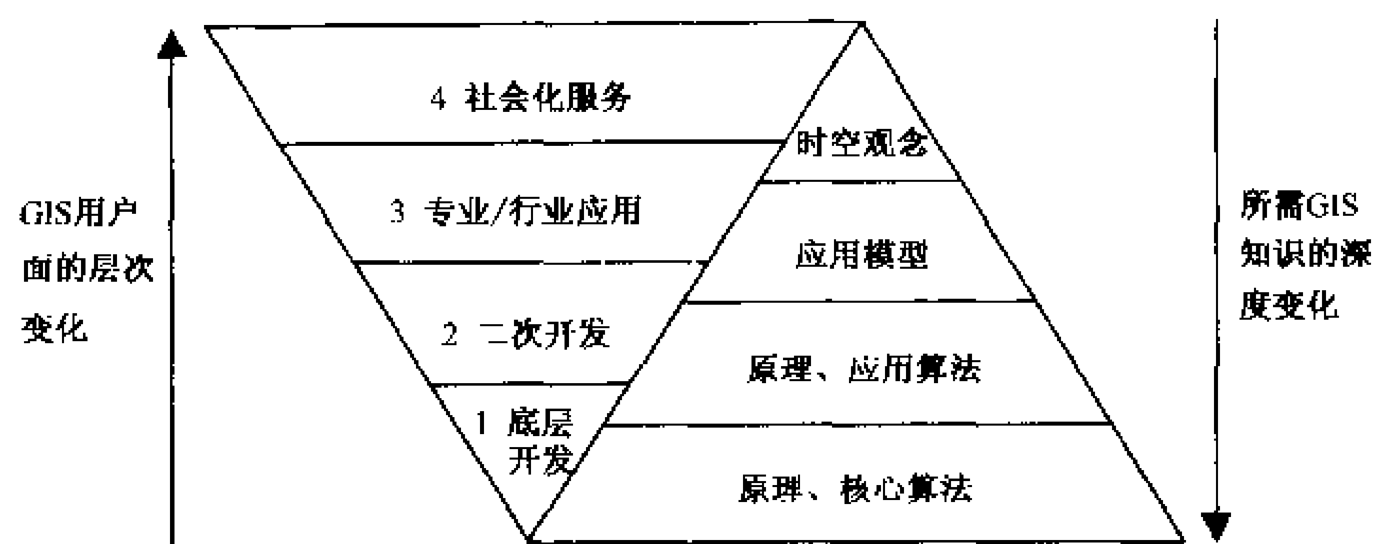
delling (C. D. Tomlin, 1990), *GIS: Principles and Applications* (D. J. Maguire, M. F. Goodchild and D. W. Rhind, 1991), *Time in GIS* (G. Langran, 1992), *GISs: Spatial Modelling and Policy Evaluation* (M. M. Fischer and P. Nijkamp, 1992), *GIS: A Computing Perspective* (M. F. Worboys, 1995), *Fundamentals of GISs* (M. N. DeMers, 1996), *Integrated Modelling for 3D GIS* (M Pilouk, 1996), *Geographic Objects with Indeterminate Boundaries* (P. A. Burrough and A. U. Frank, 1996), *Algorithmic Foundations of GIS* (M. V. Kreveld et al., 1997), *Getting Started With GISs* (K. C. Clarke, 1997), *Principles of GISs* (P. A. Burrough and R. A. McDonnell, 1998), *A Introduction to The Theory of Spatial Object Modelling* (M. Molenaar, 1998), *Geographical Information Systems* (P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind, 1999), *Fundamentals of GISs* (M. N. DeMers, 2000), *Spatial Data Quality* (W. Z. Shi, P. F. Fisher and M. F. Goodchild, 2002), *Temporal GIS: Advanced Functions for Field-Based Applications* (G. Christakos et al., 2002) 等。国内在第一本 GIS 教材《地理信息系统导论》(黄杏元等，1989) 出版使用以来，迄今已有数十本 GIS 教材或著作出版，其中不乏优秀之作，已经而且继续在教学、研究、开发与应用方面发挥积极作用。

多年来，作者在从事 GIS 方向研究生培养与 GIS 核心教学过程中，在 MGIS、Geo-Mo^{3D}和 Cyber Hong Kong 等技术系统的研究开发过程中，以及在与有关公司、企业的 GIS 研发人员的合作交流过程中，深深感受到一个问题，即研究生不同于本科生，GIS 高级研发人员不同于一般程序员，他们在 GIS、数学、计算机技术、软件开发等方面已经具备了一定的知识和经验。他们在研究生学习或产品研发（尤其是 GIS 底层开发）阶段所缺的主要是 GIS 的基本原理和核心算法方面的知识。而事实上，国内一直缺乏一本适合 GIS 研究生和 GIS 研发人员这一特定群体的教材或著作，国际上也没有特别针对这一高层次需求的专门书籍。往往为了解某一个算法、编制某一个程序，而不得不查找大量的有关资料，结果还可能一无所获。

因此，作者深感撰写出版一部关于 GIS 基本原理和核心算法的书籍的必要性和紧迫性。但由于时间关系，以及知识积累与资料准备等原因，多年来一直未能如愿。终于，最近我们在教育部“高校青年教师奖”专项基金和香港特别行政区政府有关课题的支持下，能专心在香港理工大学和中国矿业大学（北京校区）之间开展合作，将国内外有关 GIS 基本原理和算法两方面的成果进行了系统的整理、梳理、分类和总结，进而进行剖析、比较、提炼、扩展、举例和评述，并结合作者的两个研究小组近年在三维 GIS、矿山 GIS、空间数据质量与不确定性、空间数据挖掘、数字矿山关键理论与技术等方面的研究成果进行总结分析和重点介绍，完成了这部书。书中，不乏作者多年从事“GIS 原理”、“GIS 原理与算法”、“三维地学模拟理论与方法”、“三维地学模拟技术”、“土地与地理信息系统设计与开发”、“高级地理信息系统”等研究生课程教学的心得和素材积累。

GIS 的用户层结构与所需 GIS 知识结构的关系可以用下图描述：用户层越高，其涉及面越宽，所需 GIS 知识越浅；反之，用户层越低，其频及面越窄，所需 GIS 知识越深。本书的目标，就是要满足第 1～2 层次的人员对 GIS 原理与算法的需要，同时扩充

第3层次人员的GIS知识结构。



GIS用户结构与所需GIS知识结构的层次关系

完稿之日，顿觉一身轻松，总算了却一桩夙愿。希望本书能实现它的目标，同时对我国GIS教育、理论与技术发展有所促进。

目　录

下篇　GIS算法

CONTENTS

Part Ⅱ ALGORITHMS OF GIS

上篇　GIS 原理

第一章　地理认知理论与模型

认知（cognition）是指人类认识和感知其生存与生活时所经历的各个过程的总称，包括感受、发现、识别、想象、判断、记忆和学习等，是“信息获取、存储、转换、分析和利用的过程”。简言之，认知就是人类自身心智“对信息的处理过程”。人脑实质上就是一个比计算机更为复杂、更为高级的加工系统，人脑对外界的知觉、记忆、思维等一系列认知过程，可以看成是对信息的产生、接收和传递过程。

地理空间认知，是研究人类如何认识自己赖以生存的地理环境（主要指地球的四大圈层：岩石圈、水圈、大气圈和生物圈），包括位置、分布、关系、变化和规律等。人类对地理空间的认知包括地理感知、表象再现、地理记忆和地理思维四个过程。通常，地理空间认知是借助图像或地图来实现的，这里有两个重要概念，即心象地图和认知制图。心象地图是指人类对地理空间环境的一种印象或心理表征，是在过去对该地理空间环境的多次感知的基础上（包括实地考察、文献阅读、地图使用等）综合形成的，是间接的和概括的，具有不完整性、变形性、相似性、差异性。认知制图通常发生在人类使用地图的过程中，人们把新近获得的信息与地图信息综合起来进行决策，如定位、定向、寻路、导航等。研究表明，人类认知制图的功能过程是可以用计算机进行模拟的，模拟的正确程度取决于模型的合理性和输入数据的客观真实性（王家耀等，2000）。

1995 年，美国国家地理信息与分析中心（National Center for Geographical Information and Analysis，NCGIA）发表的“Advancing Geographic Information Science”报告中提出，地理信息科学的战略领域有三个，其中之一是地理空间的认知模型（Cognitive Models of Geographic Space）。1996 年，美国地理信息科学大学研究会（University Consortium for Geographic Information Science）发表的“Research Priorities for Geographic Information Science”中，把地理信息的认知（Cognition of Geographic Information）列为第二个重要议题。由此可见，地理认知理论已成为地理信息科学的重要基础理论问题。

§1.1　GIS 与广义地理系统

空间与时间是自然万物与社会现象的两个基本参照系，任何事物、任何现象都离不开这两个基本参照系。换言之，空间坐标与时间刻度是标识自然万物与社会现象的身份证。而地理系统，正是指在某一个特定的时间和特定的空间内，由两个以上的相互区别又相互联系、相互制约的地理要素或过程组成的，并具有特定功能和行为，与外界环境相互作用，并能自动调节和具有自组织功能的整体。地理系统可以用数学形式描述为：

$S=\{\Omega, T, R\}$。其中 $\Omega=\{x_1, x_2, x_3, \cdots, x_n\}$ 表示系统中 n 个不同类型地理要素的集合：其中 $T=\{t_1, t_2, t_3, \cdots, t_m\}$ 表示系统所处的 m 个不同时间（时刻或时段）；$R \subset x_1$，x_2，x_3，…，x_n 表示系统中各地理要素之间的相互联系与相互制约关系。

1.1.1 学科融合与 GIS 内涵发展

地理系统主要涉及地球表层空间。按照层次划分，可以分为岩石圈、水圈、生物圈、大气圈和电离层，它们之间在空间上有交叉。传统的 GIS 所涉及的范围主要在岩石圈和大气圈之间，即地球的表面，也就是传统地理学所研究的范畴。而自 20 世纪 70 年代以来，随着航天技术的迅猛发展，全球定位系统（Global Positioning System，GPS）、摄影测量（Photogrammetry）和遥感技术（Remote Sensing，RS）所获取的数据逐渐成为 GIS（Geographical Information System）的重要数据源。与此同时，随着计算机技术和 GIS 技术的迅猛发展，GIS 在气象、地学等涉及全球变化或地下工程（如矿

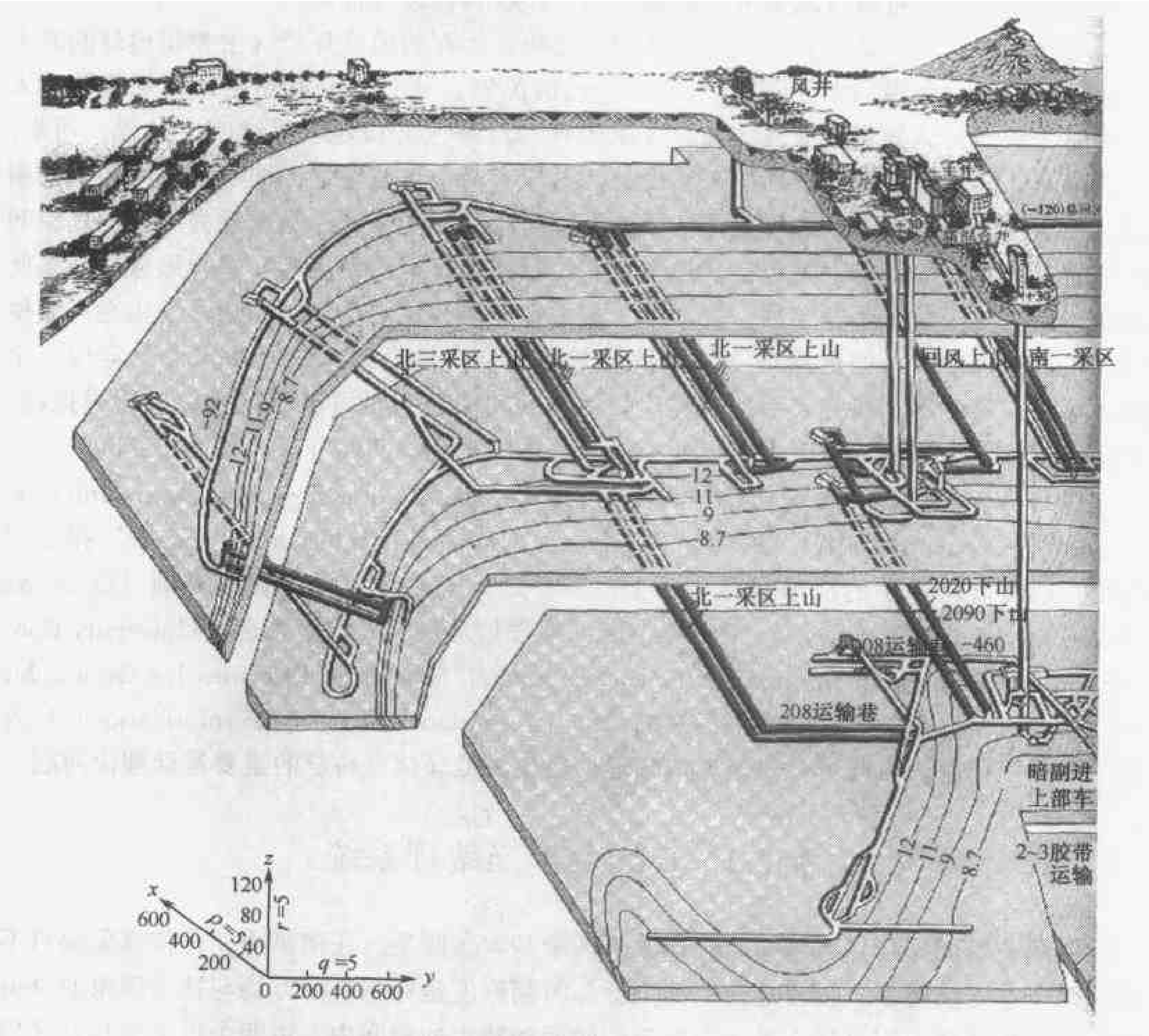


图 1-1 河北某省煤矿的井上下三维对照示意图

（陈炎光等，1992）

山、地质、隧道）领域的应用不断扩展，GIS所涉及的范围已经向上进入大气层及地球的外层空间（包括电离层），向下进入地球内部（岩石圈层之内，地表以下0～10 km），以矿山为例，如图1-1所示。

显然，这已经远远超出传统地理学和地理信息系统的研究范畴。1992年，Goodchild提出了地理信息科学（Geographical Information Science）的概念，认为GIS已不仅仅是一个技术实现，而是与计算机技术、地理学、测绘学密切相关的一门科学。钱学森（1994）、马蔼乃（1996，2003）也曾精辟地指出：地理科学是自然科学与社会科学的桥梁，地理信息科学这门以高新技术武装的技术科学的发展，带动了整个地理科学的建立与发展；地理信息科学的主要内容就是天地信息一体化网络系统，是信息社会的重要支柱之一。

学科的融合和交叉极大地推动了GIS的发展，丰富了GIS的学科内涵，包括地理空间信息认知、获取、表达、处理、共享、可视化、传输、使用等的基础理论与技术方法。人们开始重新思考和审查GIS这一多年沿用的名词（邬伦等，2001），主要表现在两个方面：

（1）Geoinformatics 和 Geomatics 的出现。这是英文词根（Geo-）和信息学单词（informatics）或其后缀（-matics）组合，两者意义相近，均表示为与地球有关的空间信息科学，包括地理学、测量学和制图学等。但又各有侧重，Geoinformatics 侧重于地球空间信息的处理、分析、表达、分发和利用，一般译为地球空间信息学或地球空间信息科学（有时也写作 Geo-information Science）；Geomatics 侧重于地球空间信息的获取、管理与绘图，可译为地球空间信息科学、地球测量学或测绘学。加拿大 *Geomatica* 杂志扉页将 Geomatics 定义为："Geomatics is a field of activities which, using a systemic approach, integrates all means used to acquire and manage spatial data required as part of scientific, administrative, legal and technical operations involved in the process of the production and management of information."

（2）以 Science 代替 System。著名的国际GIS杂志（*Int. J. Geographical Information System*）改名为 *Int. J. Geographical Information Science*，中国海外GIS协会（CPGIS）的会刊名为 *Geographic information Science*（创刊于1995年），中国的《地理学与国土研究》自2003年起改名为《地理与地理信息科学》，其英文刊名即为 *Geography And Geo-information Science*；原《武汉测绘科学大学学报》（现《武汉大学学报·信息科学版》）的英文刊名为 *Geomatics and Information Science of Wuhan University*。

由于以上发展和变化，国际上出现了一些以 Geoinformatics 命名的国际杂志，如荷兰ITC的 *Geoinfomatics*（1998年创刊）、法国的 *Geoinformatica*（1997年创刊）、亚洲遥感研究信息网（ARSRIN）与亚洲遥感协会的 *Asian Journal of Geoinformatics*（2001年创刊，泰国AIT）。国际上的一些有关机构也纷纷改名，如香港理工大学的土地测量与地理信息系的英文名称为：Land Surveying & Geo-informatics，加拿大 Laval 大学、Carlgary 大学、澳大利亚新南威尔士大学的测量专业则使用 Geomatics 命名；而加拿大的原测量署（Sector of Surveying and Mapping）则改名为 Canada Institute of Geomatics，其官方刊物更名为 *Geomatica*。中国国家基础信息中心的英文名称为 *National Geomatics Center of China*，中国矿业大学（北京校区）的测绘与土地科学系的英文名称也以

Geomatics 进行表述。

1.1.2 广义地理系统与广义地理目标

从学科融合与交叉的角度考虑，我们把涉及地球内部、表面及外部的地理系统称为广义地理系统。广义地理系统基本涵盖了当今人类生存、生活、研究和利用的现实世界整体。它所包含的各类目标称为广义地理目标。人们对广义地理目标的认识过程首先是观察，根据观察视点、尺度特征和兴趣点的不同，可以将广义地理目标分成4种基本类型。

1. 点状目标

点状目标是指可以用一对空间坐标、一个标识符和若干描述项共同表示的地理目标。包括测量控制点、电线杆、水井、水塔、矿井口、烟囱、单颗树、震中、火山口、山隘、山峰等。此外，在小比例尺情况下，某些具有面状分布特征的地理目标可以忽略其面状几何特征而归化成点，小至车站、工厂、学校、医院、机关、交通枢纽、矿点等，大至乡村、城镇、矿床等。

2. 线状目标

线状目标是指可以用一组空间坐标、一个标识符和若干描述项共同表示的地理目标，可以是直线也可以是曲线。包括各类边界线、境界线、电力线、通讯线、海岸线和抽象的等高线，以及地下管网、巷道、河流、铁路、公路等。而实际上，河流、铁路、公路、地下巷道等是有宽度的，是一个狭长的地理区域。因此在大比例尺情况下，这些狭长的地理目标的宽度是不可忽视的，要表示为双线。

3. 面状目标

面状目标是指可以用一组空间坐标或若干线状目标、一个标识符和若干描述项共同表示的地理目标，可以是平面也可以是曲面。包括地块、森林、草场、湖泊、沙漠、建筑物、乡村、矿区、城市、区县等。多数面状目标具有明确的地理边界，如建筑物、湖泊、小区、城市、矿床等。而某些面状目标并没有客观存在的、明确的边界，其边界是人为划定，如矿区、土壤类型、森林边界、草场边界、地块等。

4. 体状目标

体状地理目标是指可以用一组空间坐标或若干面状目标、一个标识符和若干描述项共同表示的地理目标，可以是规则的也可以是不规则的。体状目标均存在三维包络边界、有一定的体积（或容积）。严格意义上讲，现实世界中的绝大多数目标均属于体状目标，如土壤、矸石山、水体、云、树木、建筑物、地下巷道、地铁站、溶洞、矿体、岩层等。在二维平面的地图上，人们一般将体状目标向某一方向投影到一个平面上，用其投影的边界区域进行表示，即将体状目标描述为面状目标。

§1.2 地理抽象过程

由于地理目标的复杂性和人们认知地理目标的角度、背景、兴趣和方法的差异性，会导致对同一地理目标的不同理解和抽象，进而导致不同地理信息系统之间信息共享与互操作障碍。为了清除这一障碍，必须建立开放式的、人们共同认可的、观点统一的地理空间认知模型。目前，代表性的研究有开放地理信息系统协会（OpenGIS Consortium, OGC）（其中 OpenGIS 为 Open Geodata Interoperation Specification，开放的地理数据互操作规范）的九层次地理抽象、国际标准化组织地理信息标准化委员会（ISO-TC211）的概念抽象和中国人民解放军信息工程大学的三层次地理抽象。

OGC 将基本地理空间认知过程抽象为九个层次，如图 1-2 所示。

（1）现实世界（real world）：指实际存在的、复杂混沌的大千世界；

（2）概念世界（conceptual world）：由人类按自然语言命名的事物组成的世界；

（3）地理空间世界（geospatial world）：用 GIS 语言和地图符号表达概念世界后的结果；

（4）尺度世界（dimensional world）：经过量测确定了几何特征和定位精度的地理空间世界；

（5）项目世界（project world）：是按语义构造的尺度世界的一个可选部分，如 GIS 中的专题层（theme）；

（6）地理点列（point）：指在一个特定的地理数据层中定义的点列，与软件系统有关；

（7）几何特征（geometry）：基于点列构造的地理要素的集合特征，也与软件系统有关；

（8）地理要素（feature）：地理要素由几何特征、属性特征及空间参考系统组成，为地理信息处理提供一个参考界面；

（9）地理要素集合（feature collections）：由多个地理要素集合而成。

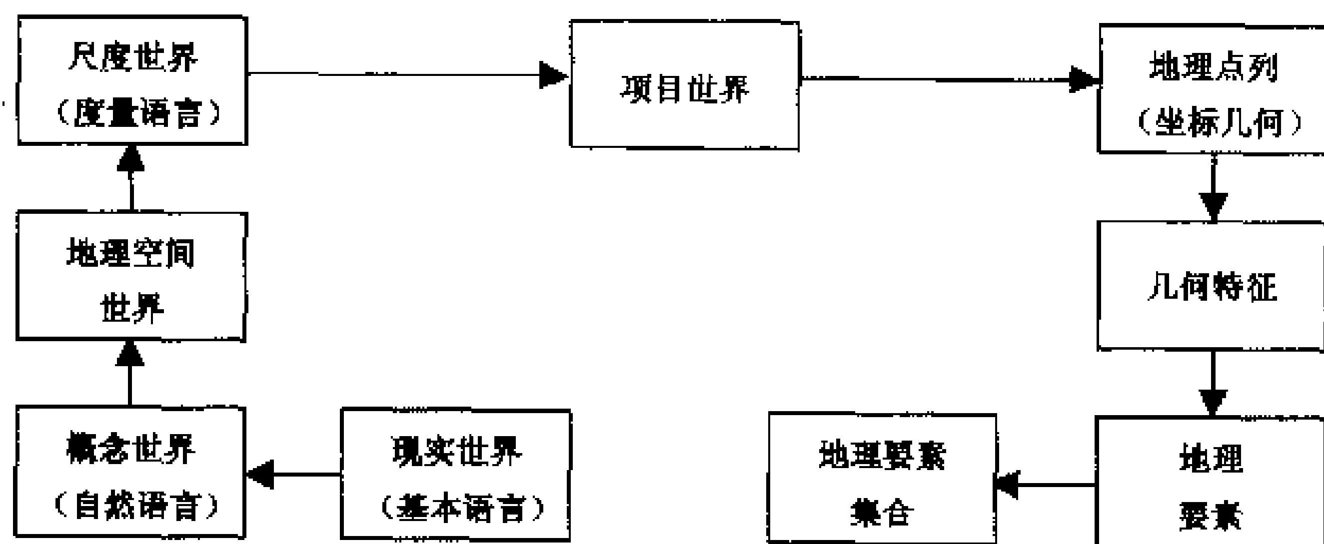


图 1-2　Open GIS 的地理抽象九层次

以上九个层次相互关联，后一层次是前一层次的派生结果。其中，前五个层次（现实、概念、地理空间、尺度和项目世界）均是对现实世界的抽象，也称感知世界，不进

行软件建模；后四个层次（地理点列、几何特征、地理要素和地理要素集合世界）均是对现实世界的数学和符号描述，也称GIS工程世界，可以进行软件建模。

ISO-TC211为了促进地理信息的互操作性和一致使用，也在制定地理空间认知的概念模式。其中规定：以数据管理和数据交换为目标，用地理信息的基本语义和结构描述地理信息，进而规范地理数据的管理，促进人们对地理空间信息的统一认识。该项工作的基本思路是分三步走：①首先，确定地理空间的论域：地理空间论域是指人们对真实世界或假想世界中感兴趣事物的讨论范围，包括地理要素的空间属性、非空间属性、功能及要素之间的相互关系等。②其次，建立概念模型：概念建模是建立真实世界整体或其局部的抽象描述或相关概念集合的过程，如以点、线、面描述地理要素形状的几何集合就是一种概念模型。③最后，构造实现模式：为了方便人们认知又适合计算机解译和处理，需要使用一种人与计算机可以共同接受的语法分析规范语言，称为概念模式语言。ISO-TC211的对地理世界的抽象过程如图1-3所示。

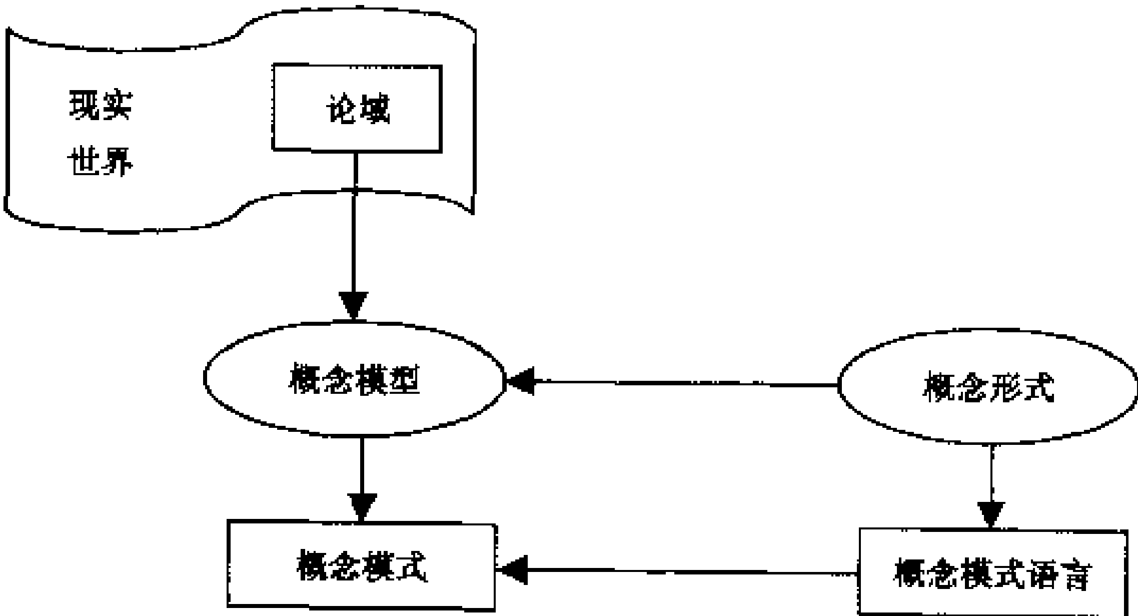


图1-3 ISO-TC211的地理抽象过程

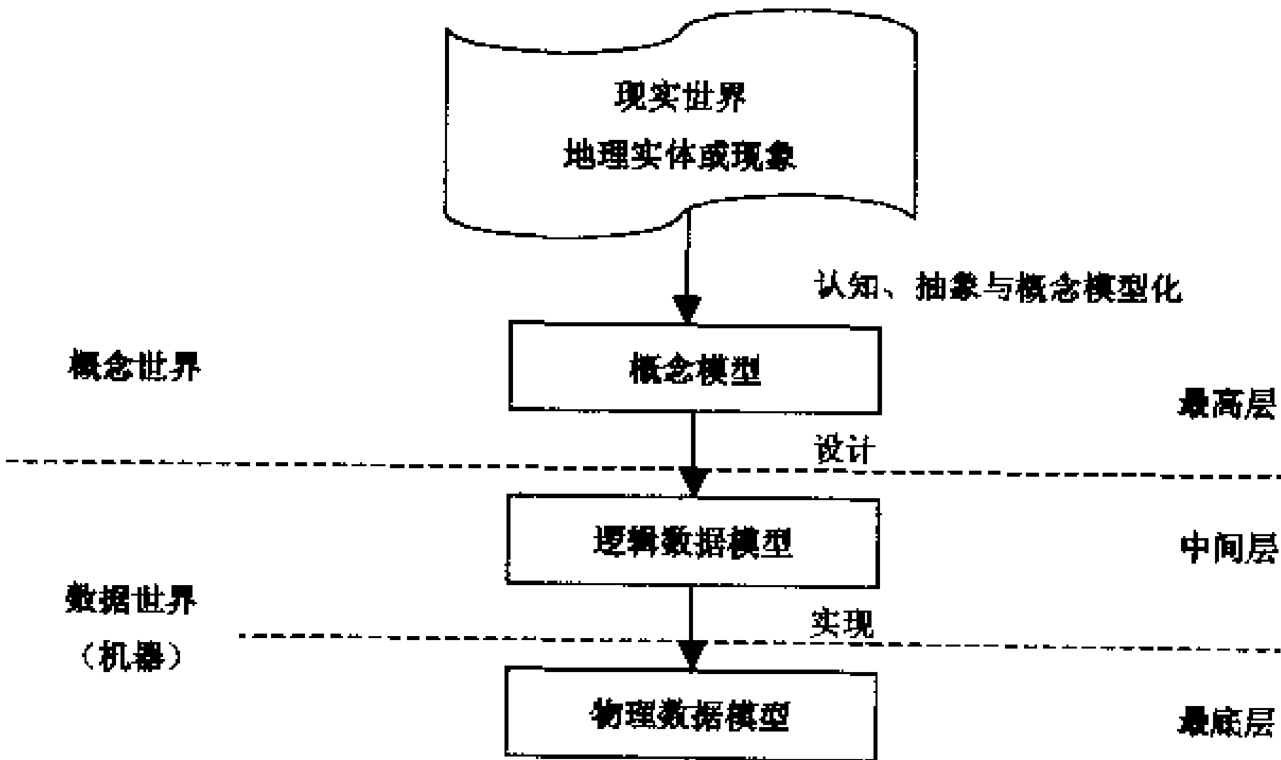


图1-4 地理抽象的三个层次
（王家耀，2001）

王家耀等（2001）针对OGC地理认知层次的繁琐和ISO-TC211抽象过程的不完整性，提出如图1-4所示的三层次模式来进行地理抽象。

其中概念模型是地理空间中实体与现象的概念集，是地理数据的语义解释，称之为地理空间认知模型。从计算机系统的角度看，它是抽象的最高层。构造概念模型应遵循以下原则：语义表达能力强；便于用户理解；独立于具体的计算实现；尽量与系统的逻辑数据模型保持统一的表达形式，而不需要或很容易向逻辑数据模型转换。逻辑数据模型是系统（如GIS）对地理数据进行表示的逻辑结构，是系统抽象的中间层，由概念模型转换而来，采用不同的实现方法具体地表达数据项、记录项等之间的关系。它是用户通过系统看到的现实世界的地理空间。逻辑模型的建立既要考虑到便于用户理解，又要考虑便于物理实现，易于转换成物理数据模型。物理数据模型则是概念模型在计算机内部具体的存储形式和操作机制，是系统抽象的最低层。

§1.3 地理认知模型

根据GIS数据组织和处理方式，目前地理空间认知模型大体上分为三类，即基于对象（object-based）、基于网络（network-based）和基于域（field-based）的认知模型（Christopher B.J.，1997），如图1-5所示。

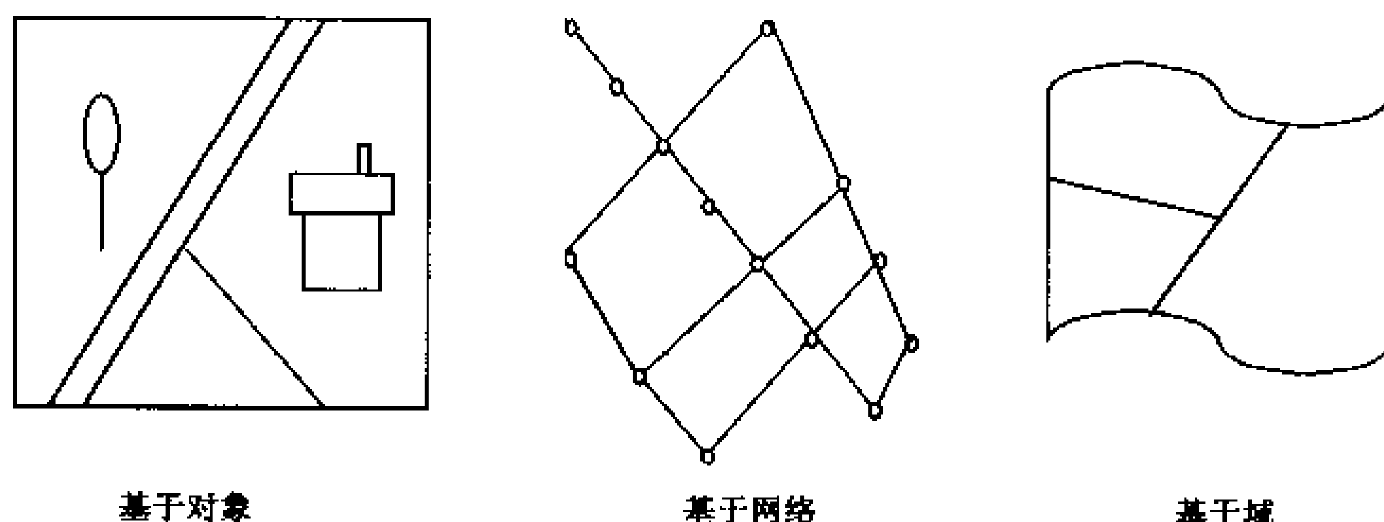


图1-5 地理空间的三种认知模型

1. 基于对象的地理认知模型

基于对象（object-based）的地理认知模型的特点是首先将整个地理空间视为一个空域，然后将各类地理实体和地理目标作为独立的对象分布在该空域中。点、线、面为其三类基本对象，基本对象之间的组合可以构成复杂对象；对象具有特定的属性，对象与对象之间保持特定的关系，如点、线、面之间的拓扑关系。单个地理目标可以基于对象的概念视图来进行标识，并按空间、时间和非空间属性及其与其他现象在空间、时间和语义上的关系来描述，如图1-6所示。

一个基于对象的视图尽管不受限制，但只适合于那些具有完整边界的地理目标，如建筑物、道路、公共设施和行政区域等人文现象以及湖泊、河流、岛屿和植被等自然现象，它们均可以作为单个地理目标来处理。

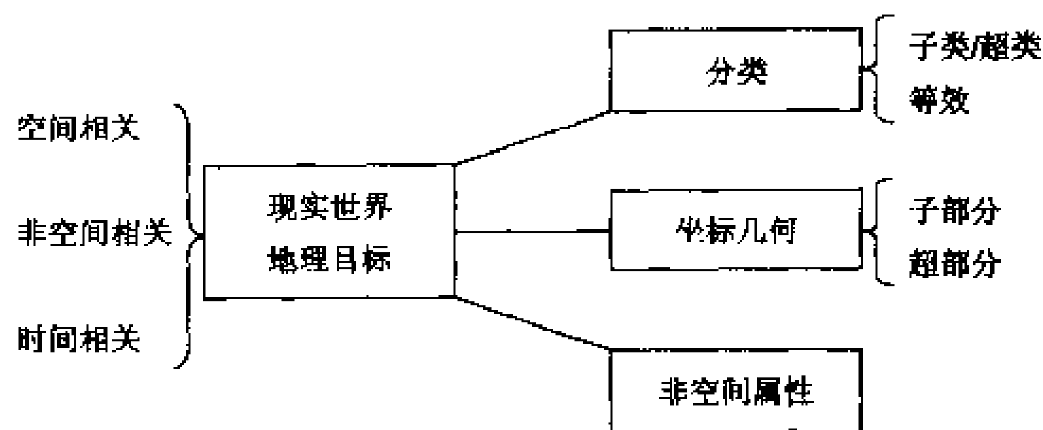


图 1-6　基于对象的单个地理目标的描述

2．基于网络的地理认知模型

基于网络（network-based）的地理认知模型与基于对象的地理认知模型有相似之处，都是描述不连续的地理目标；所不同的是网络模型需要通过目标之间的相关联接（如路径）相互连接多个地理目标。现实世界的许多地理事物和地理目标可以构成网络，如道路、地下管线、通讯线路、矿山巷道以及自然界的物质流、能量流、信息流等，均可以表示成相应点之间的连线（白玲，1997）。如果将这些点和连线分别用对象模型进行描述，网络模型就可以看成是基于点对象和线对象及其拓扑关系的集合描述。

3．基于域的地理认知模型

基于域（field-based）的地理空间认知模型的特点是把地理空间中的事物和现象作为连续的变量或体来看待，例如：大气污染、植被覆盖、土壤类型、地表温度、洋流等　域的实质是一类具有共同属性值的地理实体或地理目标的集合。根据问题的不同，域可以是 2D 的，也可以是 3D 的。如大气污染是 3D 现象，地下水渗流是 3D 现象，矿山开采沉陷也是 3D 现象。但为简单起见，这些 3D 现象也可以用 2D 域来进行描述。按表示方法的差异，可以将基于域的模型进一步区分为图斑模型、等值线模型和采样模型（据贺建忠，1995）。

1）图斑（piecewise）模型

图斑模型将一个平面划分为若干区域（region），每个区域用一个简单的数学函数来表示一种主要属性的变化。一个区域可能包含若干子区域或其他区域，但不能被其他区域分离，即区域内的任意两点之间至少存在一条完全包含在本区域内的连接路径。根据地理目标的不同，属性描述函数可以是以下之一：

（1）常量：这是简单情况，每个区域中的某种属性值保持常数，如土壤类型、植被类型和土地利用状况等地理目标的描述。

（2）线性函数：这是较复杂情况，每个区域内的某种属性值的变化是一个线性函数，要求属性变量的测量尺度是连续的。常用的不规则三角网（triangular network，TIN）就是一类典型的线性图斑模型，每一个三角面内的属性值（包括高程）的变化均是连续的和线性的。

（3）高阶函数：某些情况下，区域的属性变化很复杂，要求用高阶函数进行属性变

化描述，以便提高描述的精度。例如，可以使用二次多项式曲面来描述地形起伏的连续变化，以消除 TIN 模型中相邻三角面之间在顶点处的不连续性。

2）等值线（contour）模型

该模型用一组等间距的等值线（C_i, $i = 1, 2, \cdots\cdots, n$）将地理区域划分成一些环状区域。每个区域中的属性值介于相邻的两条等值线的属性值之间，可以沿等值线组的梯度方向进行连续插值以求其属性。

3）采样（sampling）模型

许多属性变量的值具有离散性特点，可以通过采集有限个样本点的属性值来表达区域的属性变化。根据样本点的分布密度和分布形态，可以选择格网（grids）模型或 TIN 模型来进行数据内插，进而模拟区域属性的连续变化。

参 考 文 献

白　玲．1997．GIS 中地理网络分析的研究与实践(博士学位论文)．解放军测绘学院
陈炎光，王玉浚．1992．中国煤矿开拓系统图集．徐州：中国矿业大学出版社．104
贺建忠．1995．面向对象的 GIS 空间数据模型的研究(博士学位论文)．北京大学
马蔼乃．1996．论地理科学的发展．北京大学学报(自然科学版)．32(1)：120～129
马蔼乃．2003．论地理科学．地理与地理信息科学．19(1)：1～4
钱学森．1994．论地理科学．浙江教育出版社．1～325
王家耀，陈毓芬．2000．理论地图学．北京：解放军出版社．35～65
王家耀．2001．空间信息系统原理．北京：科学出版社．70～106
邬　伦等．2001．地理信息系统——原理、方法与应用．北京：科学出版社．420～422
Christopher B. Jones. 1997. GIS and Computer Cartography. Longman Singapore Publishers (Pte) Ltd

第二章　2D GIS 空间数据模型

数据模型（data model）是描述数据库的概念集合，包括精确描述数据、数据关系、数据语义及完整性约束条件的概念。按照王家耀（2001）的地理抽象层次模式，地理数据模型是 GIS 系统抽象的中间层，即建立 GIS 的逻辑数据模型，称之为 GIS 的空间数据模型。空间数据模型是关于现实世界中空间实体及其相互间联系的概念，是建立在对地理空间的充分认识与完整抽象的地理空间认知模型（或概念模型）的基础上，并用计算机能够识别和处理的形式化语言来定义和描述现实世界地理实体、地理现象及其相互关系，是现实世界到计算机世界的直接映射。空间数据模型为描述空间数据组织和设计空间数据库提供基本方法，是 GIS 空间数据建模的基础（邬伦，2002）。

空间数据模型的发展是与数据库技术的发展密切相关的。第一代层次与网状数据库（以 1969 年 IBM 公司研制的 IMS 为标志）带动了 GIS 层次数据模型（hierarchical model）和网络数据模型（network data model）的发展；第二代关系数据库（以 Oracle, SQL server, Sybase 为代表）带动了 GIS 关系型数据模型的发展和成熟。而面向对象的数据模型技术对数据库技术的发展产生了深远影响，成为第三代数据库系统的主要标志，进而也促进了新的 GIS 面向对象数据模型（object-oriented data model）的发展。第三代数据库系统的另一个主要特征是数据库技术与其他技术的相互结合，表现为分布式数据库、工程数据库、演绎数据库、多媒体数据库、地理数据库等新技术层出不穷，进而也带动了 GIS 超图数据模型的研究与发展。

§2.1　三种传统数据模型

基于记录的数据模型把数据库定义为多种具有固定格式的记录型，每个记录型由固定数量的域或属性构成。每个域或属性具有固定的长度，使用固定长度可以简化数据库的物理结构设计。代表性的基于记录的数据模型包括三种传统的数据模型，即层次数据模型、网络数据模型和关系型数据模型。其中，层次数据模型和网络数据模型也称为结构化逻辑数据模型，关系型数据模型也称为面向操作的逻辑数据模型。

1. 层次数据模型

层次数据模型的核心是记录和系。记录表示数据，一个记录具有多个数据域，每个数据域表示一个数据项。数据之间的关系由系来表示，系可以视为一个指针。在数据库中的记录被组织为树的集合。

以图 2-1 为例，左边是某个区域（M）的地理对象的地图表示，M 包括两个面域（P_1, P_2），面域由边围成，边有两个端点。其中 L_2 为 P_1 与 P_2 的公共边。层次数据模型是一种描述该区域地理特征的倒挂树型结构（根在上，枝叶在下），见图 2-2。

由图 2-2 可见，采用层次数据模型描述地理世界虽然层次结构清晰，但是公共边、

公共点要重复描述，即边、点每重用一次，就要相应的描述一次。显然，浪费了存储空间，不利于地理要素的简洁描述。

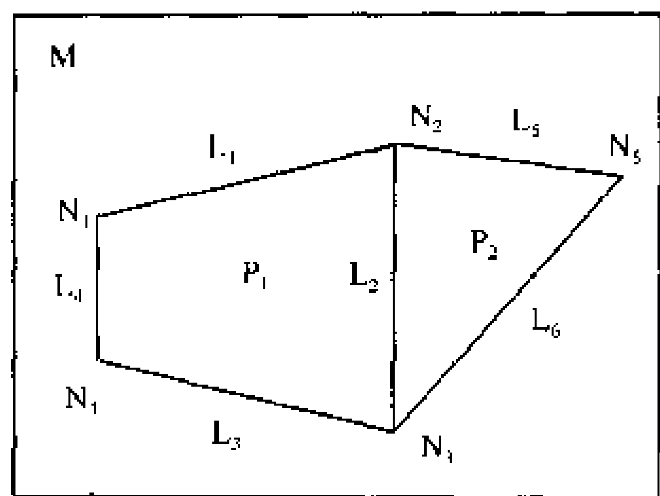


图 2-1　某区域的地图表示

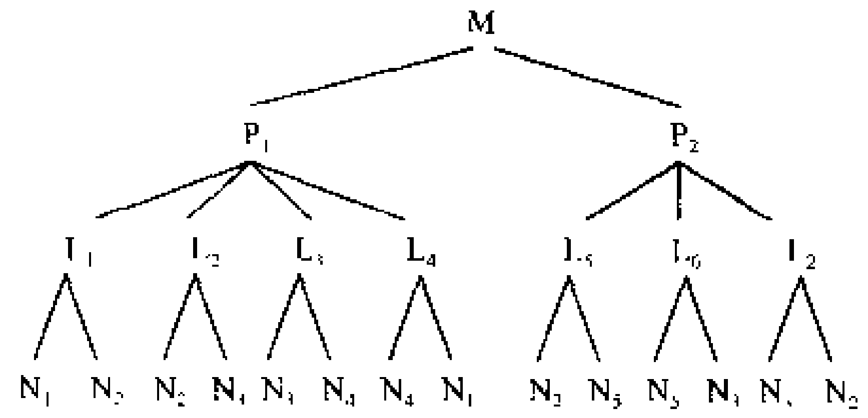


图 2-2　层次数据模型的倒挂树结构

2．网络数据模型

针对层次数据模型的重复描述的缺点，CODASYL 设计提出了网络数据模型用于网络数据库。网络数据模型类似于层次数据模型，也是用记录和系来表示数据与数据之间的关系。所不同的是数据库中的记录被组织为任意图的组合，且公共边、公共节点不必重复描述。网络数据模型描述图 2.1 区域地理要素的网络结构如图 2-3 所示。

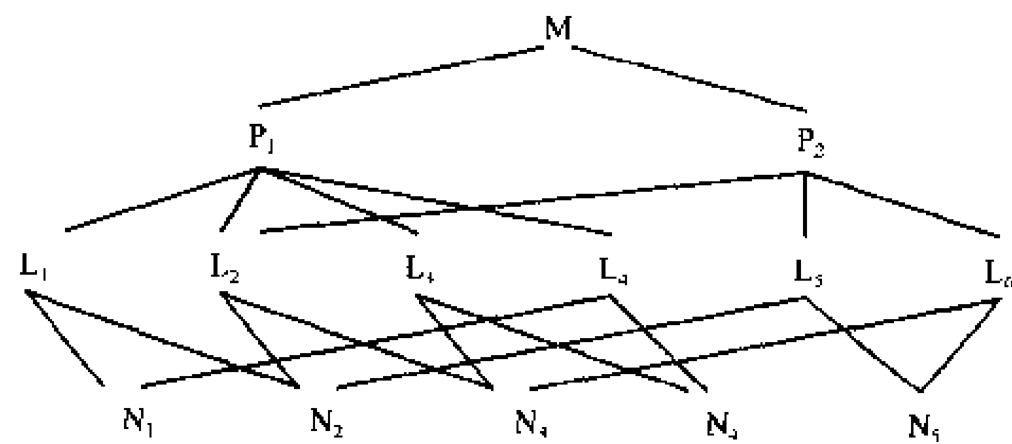


图 2-3　层次数据模型的倒挂树结构

网络数据模型的典型特征是不仅可以象层次模型一样表达一对一、一对多的关系，而且可以表达层次模型所不能表达的多对多关系，即①一个上级结点（也称父结点）可以对应多个下级结点（也称子结点）；②一个下级结点可以对应多个上级节点。

3．关系型数据模型

关系型数据模型是一种数学化的模型，它将数据的逻辑结构归结为满足一定条件的二维表或关系。一个实体由若干关系构成，而关系表的集合就构成关系模型。一组关系对应一个表；每个表有数个列；每个列具有惟一的名称，称为关系的属性；每一行数据称为一个元组，是关系表数据文件中的一条记录。

关系模型具有以下基本特征：①描述的一致性：目标之间既不用指针也不用 ID 码来联系，而是数据本身通过公共值的隐含来表达其间的关系；②操作的简单性：用关系

代数和关系运算来操作数据；③结构简单灵活：可以根据目标的复杂程度，随时调整和修改表的结构，数据更新也非常容易。

图 2-1 中的区域地理要素可以用以下 4 个关系表来描述。

表 2-1 关系 1：边界关系（E）

面域（多边形）	边界	边界长，m
P_1	L_1	4123
P_1	L_2	4000
P_1	L_3	4123
P_1	L_4	2000
P_2	L_5	3041
P_2	L_6	4610
P_2	L_2	4000

表 2-2 关系 2：边界-结点关系（E-N）

边　界	起端点（BN）	终端点（EN）
L_1	N_1	N_2
L_2	N_2	N_3
L_3	N_3	N_4
L_4	N_4	N_1
L_5	N_2	N_5
L_6	N_5	N_3

表 2-3 关系 3：结点坐标关系（C）

端　点	x	y
N_1	1000.0	4000.0
N_2	5000.0	5000.0
N_3	5000.0	1000.0
N_4	1000.0	2000.0
N_5	8000.0	4500.0

表 2-4 关系 4：面域关系（P）

面域（多边形）	面积（A，km^2）	地物类别
P_1	12.0	耕　地
P_2	6.0	开发区

目前，大多数 GIS 均采用关系型数据模型来管理属性数据，有些 GIS 如 System9 还采用关系型数据模型来管理几何数据。

§2.2 面向对象数据模型

面向对象（object-oriented）是一种方法学（paradigm），它比较自然地模拟了人类认识客观世界的方式，能够建立比较完整的、易于人们理解的软件系统的概念和机制。现实世界中无论多么复杂的实体或现象，都可以用一个对象或多个对象的集合来表达。针对软件生产质量和生产效率低下问题，采用面向对象思想发展起来的面向对象技术，是继高级语言、结构化软件设计之后的软件生产新境界，已经成为一种重要的软件系统设计和实现的软件工程方法。

1. 面向对象的基本概念

面向对象技术的核心是对象（object）和类（class）。对象是指现实世界中的实体或现象，是系统构成的基本单位。一个对象由一组属性和对这组属性进行操作的一组服务组成。属性是一些数据项，用来描述对象的静态特征；服务是一个操作序列，用来描述对象的动态特征（行为），每个对象都有一个惟一的标识号（ID）作为识别标志。类是具有部分相同属性和服务的一组对象的集合，是这些对象的统一抽象描述，其内部也包括属性和方法两个主要部分。类是对象的共性抽象，对象则是类的实例，表示为 is-instance-of 的关系；对象具有类之外的自身特有的属性和服务。类的共性抽象称为超类，类是超类的子类，表示为 is-a 的关系。一个类可能是某一类的超类，也可以是某一超类甚至是多个超类的子类。

此外，面向对象技术还有 5 个基本概念，即封装（encapsulation）、继承（inheritance）、消息（message）、多态性（polymorphism）和永久对象（permanent object）。GIS 的面向对象模型就是在上述概念的基础上，运用分类（classification）、概括（generalization）、联合（association）、聚集（aggregation）4 种语义抽象技术和继承（inheritance）与传播（propagation）2 种语义工具来完成对现实世界实体与现象的描述。

（1）分类：是把具有部分相同属性和服务的实体对象进行归类抽象的过程。

以美国地质调查局（USGS）对建筑物类的定义为例，它的基本属性有：所属街道、门牌号、房主等，这是城市 GIS 中房屋建筑类的共有属性项。

（2）概括：是把具有部分相同属性和服务的类进行进一步归类抽象为某种超类（superclass）的过程。

以矿山 GIS 为例，煤矿、铁矿、铜矿、金矿、石英矿、石膏矿等，无论所开采的矿产资源是什么，总之都是矿山。因此“矿山”就是对各类矿的进一步抽象，是一个超类，它具有各类矿所共有的“所属地理区域”、“所有制形式”、“年产量”、“投产时间”等属性项。而作为其子类的煤矿，还可以依据“所有制形式”的不同，归纳为“国有煤矿”、“县办煤矿”、“乡镇煤矿”、“个体小窑”等子子类。

（3）联合：是把一组属于同一类中的若干具有部分相同属性值的对象组合起来，形成一个新的集合对象的过程。集合对象中的个体对象称为它的成员对象，表示为 is-member-of 的关系。显然，联合不同于概括：概括是对类的进一步抽象得到超类，而联合是对类中的具体对象进行合并得到新对象。

以房产GIS为例，某一街区的高层建筑物这一类中，有N幢高层建筑进行了抗震结构设计，均采用了滑动地基、增强基础圈梁等特殊结构措施，可以把他们联合起来进行管理，建立一个新的集合对象，即抗震型高层建筑群。

（4）聚集：聚集与联合不同，它是把一组属于不同类中的若干对象组合起来，形成一个更高级别的复合对象的过程。复合对象中的个体对象称为它的组件对象，表示为is-part-of的关系。这一技术非常适合于描述复杂对象。

以城市GIS为例，东长安街、王府井大街等是街道类的对象，王府饭店、北京饭店等是宾馆类对象，北京市百货商场、新东安市场、王府井书店等是商贸类对象，这些对象聚集起来就形成了王府井街区这一新的更高级别的复合对象。

（5）继承：是服务于概括的语义工具，是从上到下作用的。超类的属性和服务可以无条件的被它的子类所继承，继承是一种强有力的建模工具，它有助于进行共享说明和应用的实现，提供了对现实世界的简明精确的描述，从而大大减少信息冗余，并有效地保持系统的一致性。

继承有单一继承，也有多方继承。所谓单一继承是指子类仅有一个超类父亲，而多方继承是指子类有多个超类父亲。多方继承的实际意义是：一个类的属性和服务可以是多个超类父亲的属性与服务的综合，GIS中经常遇到这样的问题。以城市GIS为例，如图2-4所示，城市体系包括交通、街区、绿地等主体部分，涉及轻轨、地铁、环路、公路、街道、建筑、园林、草坪、树林等基本部分。其中街道既可以行人通车，成为交通体系的一部分；也可以从事商业活动，成为商业街区的重要组成部分；还可以在街道两侧和中间隔离带搞绿化，成为城市的绿色走廊。因此，街道就成为城市GIS中一类典型的具有多方继承关系的特殊类。此外，城市园林也是如此。

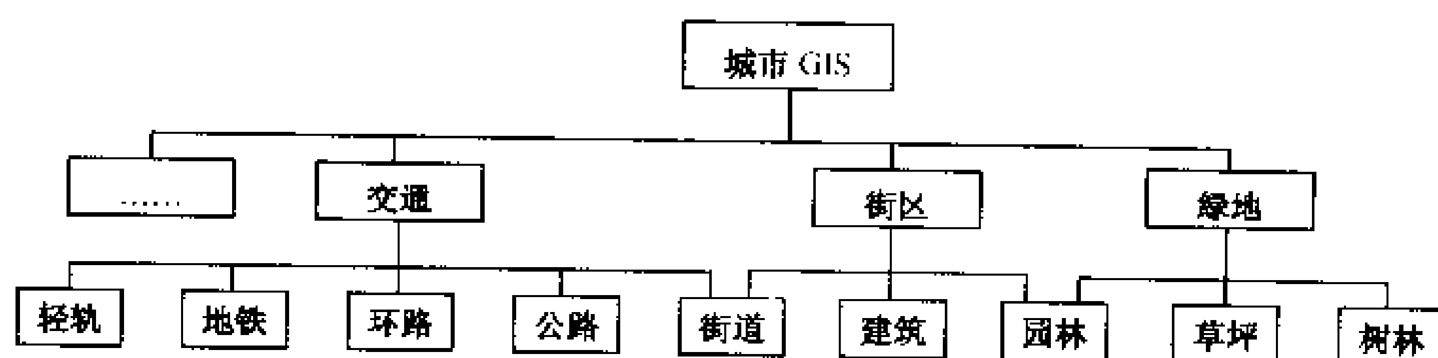


图2-4　城市GIS中街道、园林类的多方继承关系

（6）传播：是服务于联合与聚集的语义工具，是从下往上作用的。它通过强制性手段将成员对象或组件对象的相关属性和操作传递给集合对象或复合对象（通称为复杂对象）。也就是说，复杂对象的某些属性值不单独存放，而是从它的成员对象或组件对象的成员中提取或派生。所谓提取，是指直接从成员对象或组件对象中获得某种数据或信息，如多边形的坐标数据并不直接存储在其多边形数据文件中，而是存储在弧段与结点数据文件中，需要时从中提取。而所谓派生，是指由相关联的数据或信息经过某种简单的四则运算，得到一种新的数据或信息的过程。如一个县的人口可以由它所关联的乡镇的人口求和得到。

基于以上机制，当对象或组件对象的相关属性被改变后，集合对象或复合对象的属性无须专门修改，而会自动从改变后的成员对象或组件对象中传播过来。因此，传播工

具也可以极大地减少信息冗余和有效地保持系统的一致性。

需要指出的是，继承和传播作为面向对象数据模型的两种重要的语义工具，其功能是相似的，但其服务对象、作用方式、作用内容却不一样：① 继承是服务于概括，对父类和子类发生作用；传播是服务于联合和聚集，对复杂对象与简单对象（成员对象或组件对象）发生作用；② 继承是隐含的，由系统提供内部机制，只需声明父类与子类的关系；传播是强制性的，需要在复杂对象中显式定义其对应的简单对象，并申明传播哪些属性，以及如何传播；③ 继承包括了属性和服务；而传播一般仅涉及属性。

2. 面向对象的建模实例

以一个矿山 GIS 的面向对象建模为例。某一矿井位于东部平原地区，其主要地貌特征和地理要素有树林、草坪、农田、池塘、晒场等；地面建筑密集，地面除矿山的办公楼、矿灯房、澡塘、机修车间、木材库、选煤楼、洗煤厂等各类建筑设施之外，还有村庄民房，灌溉机井等；地面道路体系有矿区运煤铁路、公路、田间机车道等；地下管线有供水管网、排水管网、热力管网、煤气管网等；自地表以下至煤系地层底板涉及第

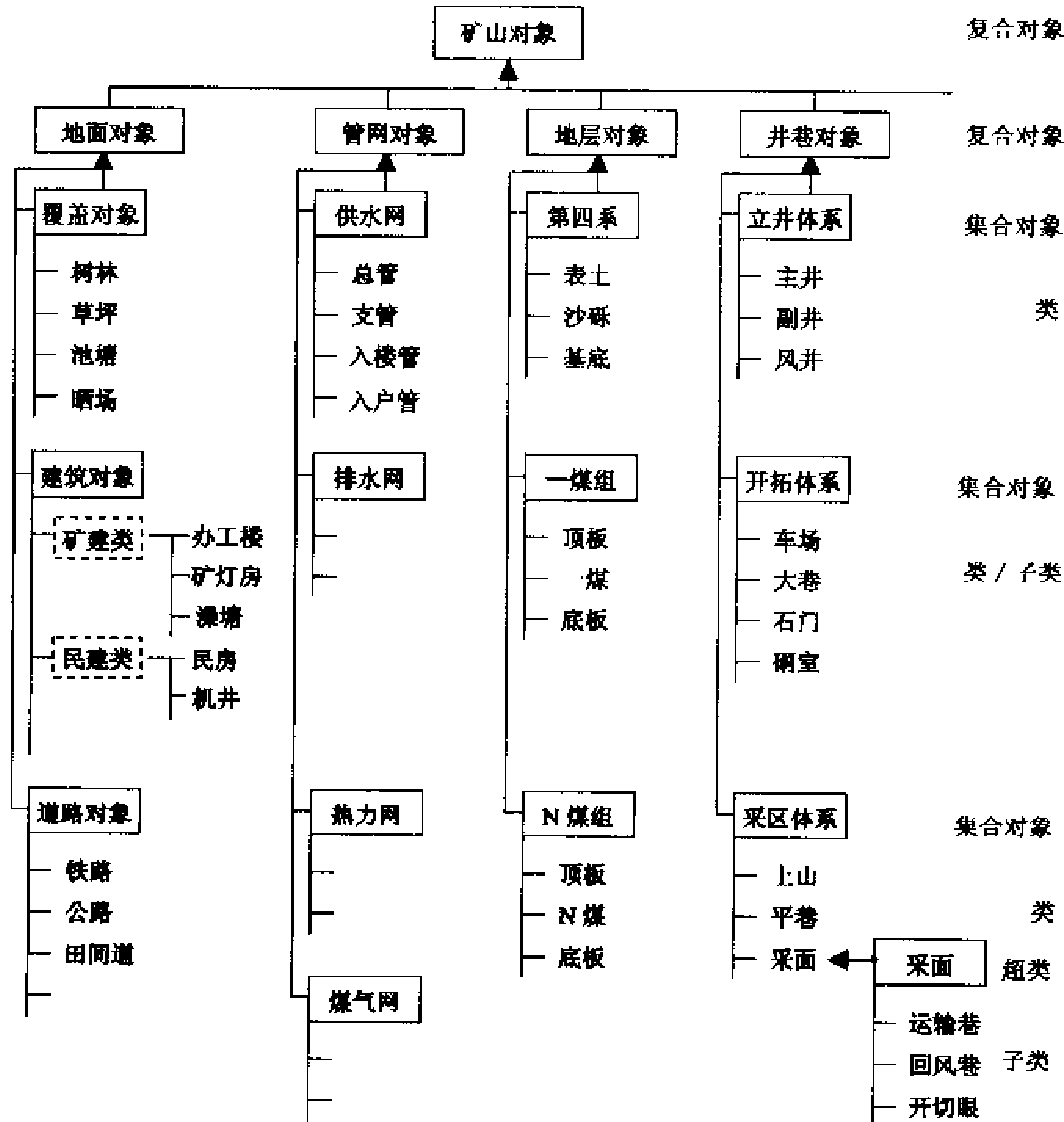


图 2-5 矿山实体的面向对象数据模型体系

四系、二叠系、寒武系、奥陶系表土层、流沙层、砾石层、砂岩、泥岩、页岩、灰岩、煤等岩层；矿井为主副井形式，分两个水平开采，当前生产水平为第一水平，两翼开拓，各分3个采区，每个采区内设计有5～8个开采工作面不等；每个工作面由运输巷、回风巷、开切眼等基本工程组成……显然，这是一个相当复杂的大系统，不仅涉及到地面地理要素，还主要涉及到地下地层和矿井地下工程。要准确、完整、有效地对其进行描述和表达，是三类传统的数据模型难以做到的。

若采用面向对象的数据模型，矿山实体的面向对象建模如图2-5所示，则可以较完整方便地表达出该复杂目标体系的结构。

§2.3 空间数据组织与管理

GIS区别于一般信息系统的主要特点是：GIS中的数据具有空间分布特征，系统的分析与应用无不与地理位置直接关联。GIS按：①某一已知坐标系中的位置，②与位置无关的属性，③实体间的空间关系来描述地理实体（Peter A. B.，1986）。GIS的这一特点，决定GIS的数据组织方式、数据库设计等均不同于通用数据库管理系统。通用数据库大多是按某种既定格式的表格方式进行管理，并提供简单的数据检索和统计分析功能。而GIS则不同，它的主要目的是要为各类决策提供空间查询、空间量算与空间分析服务，这就要求GIS必须具有更强于一般信息系统的数据组织与管理能力，特别是在空间数据组织与管理方面的能力。

2.3.1 空间数据的特征

空间数据具有六个基本特征：

（1）空间特征：空间特征包括空间位置（坐标）和空间分布，这就要求GIS除了必须具备通用数据库管理系统或文件系统的关键字索引和辅助关键字索引之外，还需建立空间索引机制。

（2）非结构化特征：由于地理实体或地理现象的非结构化特征，决定GIS中空间数据的非结构化特征。如一条弧段可能只有两个坐标对，也可能有千百个坐标对，因此弧段记录的长度是不定的；此外，一个多边形可能只由一条弧段封闭而成，也可能由若干条弧段首尾相连而成，因此多边形记录是多条弧段的嵌套。这种变长记录和不定结构的要求，是一般关系型数据库所不能满足的。

（3）空间关系特征：空间数据除了要描述地理实体的空间坐标和空间分布之外，还要描述地理实体之间的空间关系以及实体组成元素之间的拓扑关系（如点与线、线与面等）。这给空间数据的一致性和完整性维护增加了困难。特别是某些几何对象并不直接记录其坐标信息，如面状目标仅记录组成它的弧段标识，因而在查找、显示和分析时均要操纵和检索多个数据文件。

（4）分类编码特征：为了惟一识别地理实体和共享空间数据，每一个地理实体均分配一个分类编码。这种编码可能是按国家标准、行业标准或地区标准进行的，也可能是全球性的。通过分类编码将空间数据和属性数据关联起来。

(5) 多尺度特征：如1.1.2节所述，由于空间认知水平、认知精度和比例尺等的不同，地理实体的表现形式也不相同。这就要求空间数据库具备有效的多尺度空间数据组织与管理功能，这也是一般关系型数据库所不具备的。

(6) 海量性特征：由于地理区域的广大性、地理数据的多源性以及空间数据分辨率的不断提高，GIS中的数据量往往要比一般事务性信息系统的数据量大得多。例如，一个城市GIS的数据量可能达几十个GB，若考虑影像数据的存储，则可能超过几百个GB。因为数据量巨大，需要在二维空间上划分为块和图幅、在垂直方向上划分为层来进行管理。

2.3.2 空间数据库管理系统

空间数据模型的发展是与数据库技术的发展密切相关的，空间数据库管理系统更是与数据库技术的发展密不可分。按发展的轨迹，可以将GIS空间数据管理系统分为文件-关系型、全关系型、对象-关系型和纯对象型4种类型。

1. 文件-关系型管理

由于空间数据的以上特征，早期关系型数据库难以满足空间数据管理的要求。因此，大部分GIS软件采用混合管理的模式，即用文件系统管理几何图形数据，用商用关系型数据库管理属性数据，两者之间通过目标标识或内部连接码进行连接，如图2-6所示。

OID（目标ID或内部连接码）	图形数据
	属性数据

图2-6 GIS的文件-关系型数据连接

在这一管理模式中，除通过OID（Object ID）连接之外，图形数据和属性数据几乎是完全独立组织、管理与检索的。其中，图形系统采用高级语言（如Fortran、C，Delphi等）编程管理，可以直接操纵数据文件，因而图形用户界面与图形文件处理是一体的，两者中间没有逻辑裂缝。但由于早期的数据库系统不提供高级语言的接口，只能采用数据库操纵语言，因此图形用户界面和属性用户界面是分开的。在GIS工作过程中，通常需要同时启动图形文件系统和关系数据库系统，甚至两个系统来回切换，使用起来很不方便，如图2-7所示。

近年来，随着数据库技术的发展，越来越多的数据库系统提供了高级语言的接口，使得GIS可以在图形环境下直接操纵属性数据，并通过高级语言的对话框和列表框显示属性数据；或通过对话框输入SQL语句，并将该语句通过高级语言与数据库的接口来查询属性数据，然后在GIS的用户界面下显示查询结果。这种工作模式，图形与属性完全在一个界面下进行查询与维护，而不需要启动一个完整的数据库管理系统，用户甚至不知道何时调用了数据库系统。

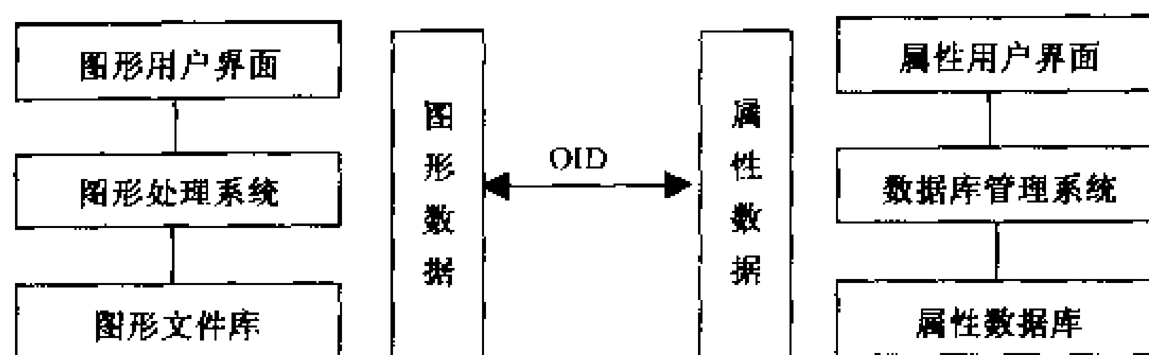


图 2-7　基于 OID 的文件-关系型图形-属性互操作

在 ODBC（Open Data Base Consortium，开放性数据库连接协议）推出之前，各数据库厂商分别提供一套自己的与高级语言的接口程序。因此，GIS 软件开发商就不得不针对每个数据库系统开发一套自己的接口程序，导致在数据共享（或数据复用）上受到限制。ODBC 推出之后，GIS 软件开发商只要开发 GIS 与 ODBC 的接口，就可以将属性数据与任何一个支持 ODBC 协议的关系型数据库管理系统连接。

无论是通过高级语言还是 ODBC 与关系型数据库连接，GIS 用户都是在同一个界面下处理图形和属性数据，如图 2-8 所示，称为混合方式。该方式要比图 2-7 的方式方便得多。

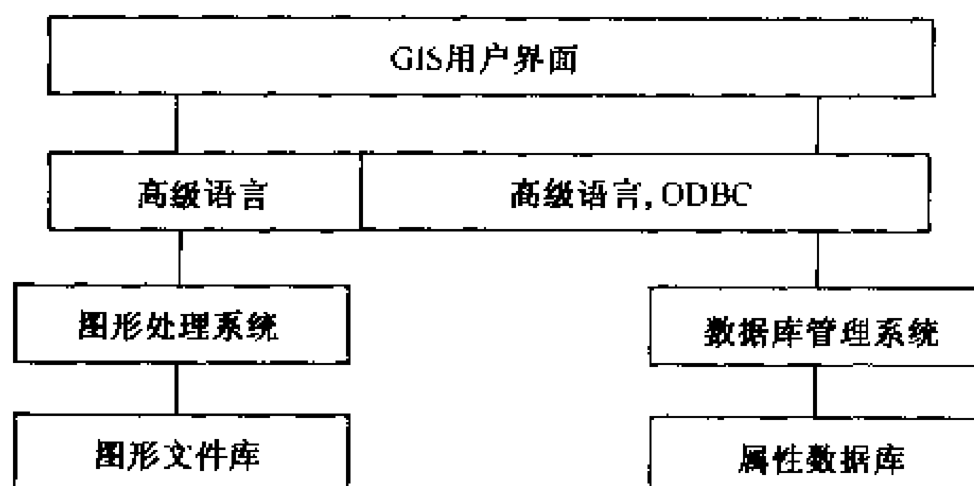


图 2-8　基于高级语言和 ODBC 的文件-关系型图形-属性互操作

采用文件-关系型模式还不能说是真正意义上的空间数据库管理系统。因为文件管理的功能较弱，不能方便地处理大区域图幅分割和地图拼接问题，特别是在数据的安全性、一致性、完整性以及数据损坏后的恢复方面缺少基本功能。在多用户操作的并行控制方面，要比商用数据库系统逊色得多。因此，许多 GIS 软件开发商一直在致力于寻找合适的商用数据库管理系统来同时管理图形数据与属性数据。

2．全关系型管理

全关系型数据库管理系统是指图形和属性数据都用某一关系数据库管理系统进行管理，GIS 软件开发商直接在某一关系型数据库管理系统的基础上进行开发，使所开发的 GIS 不仅能管理结构化的属性数据，还能管理非结构化的图形数据。

用关系型数据库系统管理图形数据有两种方式。

1）基于关系模型

该方式按2.1节的关系数据模型组织图形数据。其缺点是：由于涉及一系列关系连接运算，查询分析相当费时，效率不高。以图2-1为例，为了显示多边形P_1，需要查找组成该多边形的边界点坐标。这涉及到四个关系表即关系P，E，N和C。因此，需作多次运算，效率不高。查询语句如下：

```
Select X, Y
From P, E, N, C
Where P·P# = E·P# and E·E# = N·E#
    and(N·BN = C·N# and N·EN = C·N#)
```

2）将图形数据的变长部分处理成二进制Block字段

目前，大部分关系型数据库管理系统均提供了二进制块的字段域，以适应多媒体数据或可变长文本数据的管理。GIS于是利用这种功能，把图形的坐标数据当作一个二进制块，交给关系型数据库管理系统进行管理。这种方式虽然省去了如前所述的大量关系操作，但是二进制块的读写效率远比定长结构的属性字段慢得多，特别是涉及到对象嵌套时，速度更慢。

3．对象-关系型管理

如上所述，采用全关系型模式管理GIS空间数据效率不高，而非结构化的空间数据管理对于数据库管理系统来说又十分重要。所以许多数据库管理系统软件厂商纷纷在关系数据库系统中进行扩展，使之能直接存储和管理非结构化的空间数据，如Ingres，Informix和Oracle等都推出了用于空间数据管理的专门模块，其中定义了操纵点、线、面、圆、矩形等空间对象的API函数。由于这些API函数将各种空间对象的数据结构进行了预定义，而且一般不带拓扑关系，用户使用时必须满足它的数据结构要求，用户（即使是GIS开发商）也不能根据GIS的要求对其进行再定义。

这种基于关系扩展的对象-关系型管理模式，主要解决了空间数据的变长记录问题，效率提高了很多，已经在GIS领域获得较多应用。但是，它仍然没有解决对象嵌套的问题，空间数据结构尤其是拓扑结构不能由用户进行定义，因此使用上仍然受到较大限制。

4．纯对象型管理

由2.2节讨论可知，采用面向对象模型的纯对象型管理方式最适合空间数据的表达与管理，它不仅支持变长记录，而且支持对象嵌套、信息继承与传播。纯对象型空间数据管理系统允许用户定义对象和对象的数据结构（包括拓扑结构），以及它的操作。这种空间数据结构可以是不带拓扑关系的面条数据结构（如等高线），也可以是带拓扑关系的拓扑数据结构。当采用拓扑数据结构时，往往涉及对象的嵌套、连接和对象信息（包括属性，甚至操作）的继承与传播。

目前推出的面向对象数据库管理系统有Jasmine，O2等。同时，也推出了一些基于

纯对象型管理的 GIS，如 GDE 等。尽管如此，面向对象的数据库系统在 GIS 领域还不是很普遍，如 SmallWorld。

2.3.3 空间数据的组织

无论采用上述何种模式管理空间数据，空间数据的组织方式均非常重要。不同的管理模式所对应的空间数据组织方式可能不一样，不同的 GIS 系统之间，其空间数据组织方式也不一样。本节以文件-关系型模式为例，讨论空间数据的组织管理问题。

1. 图幅数据组织

GIS 中将某一问题域或某一项 GIS 任务称为一个 GIS 工程。随着 GIS 工程范围的扩大（如扩至全矿区、全市、全省甚至全国），必须将空间数据进行分幅管理（地图制图中的地图分幅概念的适用）。根据问题的需要和工作方便，通常将一幅图或几幅邻接的图当作一个工作单元，称之为工作区（workspace）。其组成关系如图 2-9 所示。

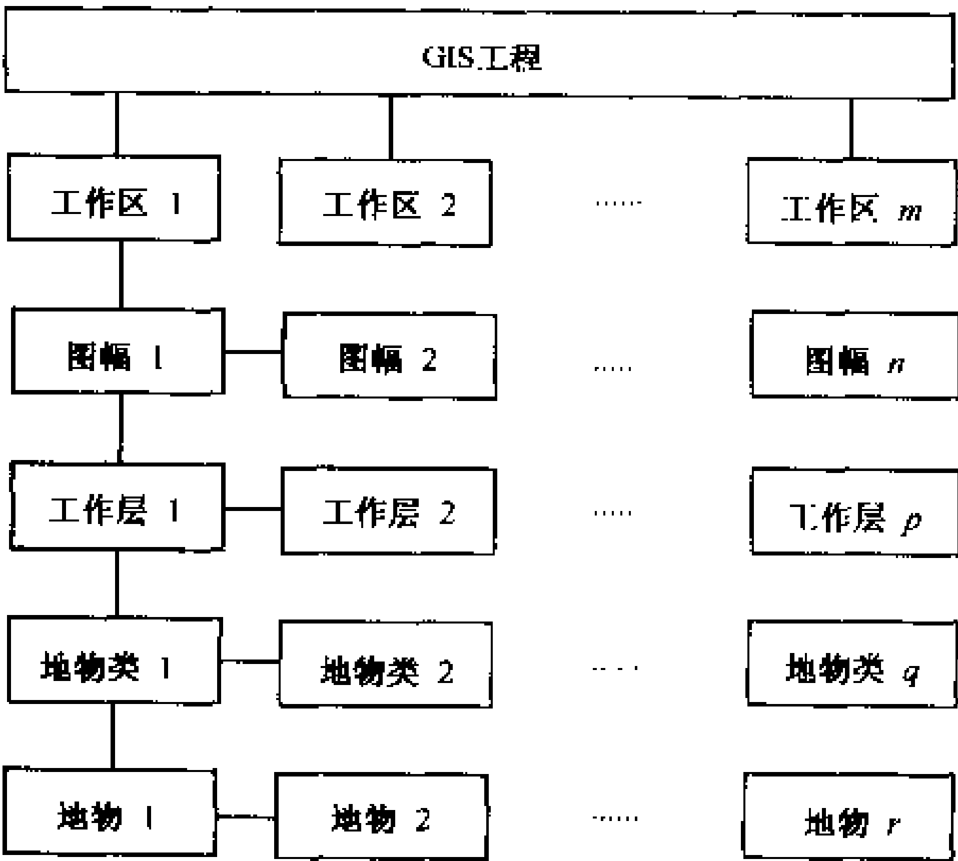


图 2-9 GIS 数据的组织管理结构

工作区中，除了包含相应图幅的各层空间数据之外，还包含对数据库的连接与操作。用户可以随时将当前工作环境以某一工作区版本的方式存储下来，下次（包括重新开机、重新启动之后）打开该工作区版本时，就可以直接恢复进入原有工作状态。

工作区由若干工作层组成，工作层在平面上可能与工作区范围一致，但在垂直方向上则因软件系统的不同而定义和叫法不同，如 ARC/INFO 的工作层称为 coverage，MGE 的工作层称为 cotalog，Titian 的工作层称为 layer。

一个工作层可以是一个逻辑层，也可以是某一覆盖层（如 layer）。工作层由一种或多种地物类（feature class）组成，用户可以根据需要任意定义。

地物类是类型相同的地物总称。同类地物一般用相同的显示颜色和符号表示，而且严格按点状地物类、线状地物类或面状地物类进行划分。有些软件将点状地物、线状地物、面状地物分不同的数据文件进行管理，如 ARC/INFO。而某些软件则将工作区中的所有地物（无论点、线、面）交由一个文件统一管理，如 MGE，Geostar 和 Titian 等。

2. 图形库管理

图形库管理即为工程管理。从 GIS 工作过程来说，每个工作区或工作层可以形成一个独立的工作单元，用户可以在其中进行各种操作。而实际上，作为一项工程而言，工作区与工作区之间、工作层与工作层之间在逻辑上和物理上都是不可分的整体，应该允许用户在整个工程内进行任意开窗、放大、漫游、查询、分析和制图。这就涉及到海量数据管理的问题，因为一项工程可能涉及到成千上万个工作区。对海量数据的管理是大型 GIS 必备的基本功能，其管理效率是衡量 GIS 优劣的重要指标之一。

工程管理一般是建立图幅索引，即在工作区范围建立二维空间索引，如图2-10所示。

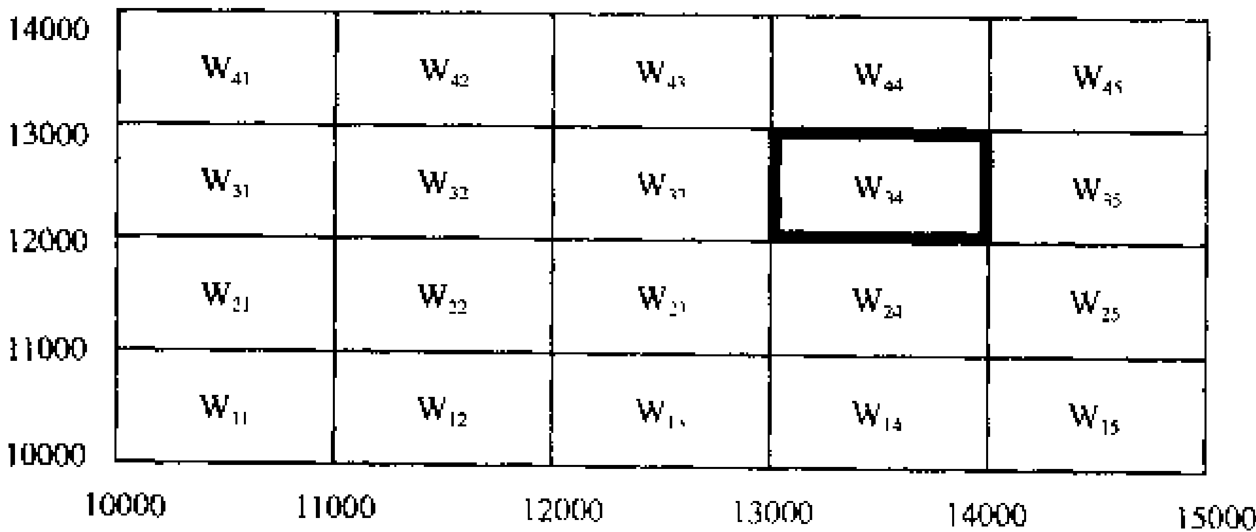


图 2-10 工作区索引
（据龚健雅，2001）

由一个索引文件记录每个工作区的范围，如 W_{34} 的范围坐标为（13000，12000，14000，13000），就可以建立工程与工作区的关系。用户根据图幅索引文件，可以在工程界面下开窗进入任一工作区或某几个工作区。

在 Intranet 和 Internet 环境下，工程除了要进行工作区管理之外，还要进行数据库并发控制管理。过去，为了数据安全起见，一般禁止多个用户同时对某一个工作区进行修改和编辑，此时工程管理要记录哪些工作区已经打开并在进行编辑，若有其他用户进入该工作区，则要立即提出警告。而在共享互动的企业级 GIS 环境下，往往需要多个用户同时对同一个工作区进行编辑和其他操作。此时，就需要将并发控制锁定在对象级而不是工作区级，只要限制多用户对同一对象的同时编辑与操作即可。如 Versioning 技术可以实现多个用户同时对一个文件的操作。

3. 属性数据组织

属性数据由关系型数据库进行管理，其文件组织方式因GIS软件而异，以下结合几个GIS软件介绍3种典型的属性数据文件的组织方法。

(1) 与工作层对应的组织方式：ARC/INFO的属性数据文件一般建立在对应的coverage目录之下。无论一个coverage包括多少地物类，仅有一个AAT表（记录属性数据）和一个PAT表（记录路径）。为了表达不同地物类的不同属性项，可以按每个地物类建立一个扩展的属性表，让它们通过地物编码和内部连接码与AAT表或PAT表相连。因此，在查询某一空间地物的属性时，先从AAT表或PAT表中得到部分信息，再通过关系连接查询到扩展信息。

(2) 与地物类对应的组织方式：MGE的属性数据文件是建立在地物类的基础上，而且所有属性文件均放在对应的工程目录之下。也就是说，不同工作区的相同地物类的属性是放在一起的，这样便于属性的工程管理，而且大大提高了在工程范围内查找某一属性的速度。需要注意的是，MGE并不要求每个地物类都带有属性表，如陡坎等无关紧要的地物就可以不要属性表，这为GIS的空间数据组织提供了一定的灵活性。

(3) 混合方式：Geostar的属性数据文件组织与管理方式吸收了前两者的优点。在Geostar中，既可以对每一个地物类设计属性表，也可以对属性项相同或相近的多个地物类设计一个公用的属性表。以交通GIS为例，高速公路、一级公路、二级公路、乡镇公路等，它们的地物类型编码可能不同，但它们的属性项可能相同，因而它们可以共有一个属性表，以便于查询、显示和最佳路径分析。Geostar的属性数据文件的组织则与MGE基本类似：在工程建立之前，属性数据文件位于与工作区平行的目录之下；在工程建立之后，则直接位于工程目录之下。一个属性文件包括了该工程内所有同类空间对象的属性。当属性文件趋于庞大时，则有必要建立关键字索引机制。

§2.4 空间索引机制

空间索引是指依据空间对象的位置和形状或空间对象之间的某种空间关系，按一定顺序排列的一种数据结构，其中包括空间对象的概要信息，如对象的标识、外接矩形及指向该空间对象实体的指针。作为一种辅助性的空间数据结构，空间索引介于空间操作算法和空间对象之间。它通过筛选作用，排除大量与特定空间操作无关的空间对象，从而提高空间操作的速度和效率。空间索引的性能优劣直接影响空间数据库和GIS的整体功能，是空间数据库和GIS的一项关键技术（邬伦，2002）。

对于传统的基于图形分幅的空间数据管理系统，可以根据鼠标所在位置的坐标快速计算当前鼠标落在那一个工作区内，这是GIS中原始的空间索引机制。这种机制在小工程条件下可以适用，而在大工程环境下，随着工作区数量的增大，其计算效率将降低。

GIS的发展已经在向无缝空间数据库的方向迈进。在空间数据的无缝管理模式下，是没有地图分幅概念的。当要查询某一个地物时，鼠标所指到底落在哪一个空间地物上，如果没有一种新的有效的空间索引机制，就需要对整个工作区中空间对象进行点

（鼠标点）与点（空间对象点）的距离比较，或进行点在线上、点在多边形内等一系列判断计算。这些计算是很复杂和费时的。因此，需要在整个空间数据库内建立空间索引。除了常用的窗坐标索引、格网索引机制之外，大型空间数据库还采用自顶向下、逐级划分的模式，代表性的有 BSP 树、KDB 树、R 树、R^+ 树和 CELL 树等索引方式。

1. 窗坐标索引机制

对象范围索引的机制是：在记录每个空间对象的坐标时，同时记录其最大最小窗坐标。所谓窗坐标，是指完全包含该空间对象的最小标准矩形（其边与坐标轴平行）的角点坐标，如图 2-11 所示。有了最大最小窗坐标的记录，在检索某一范围内的空间地物时，首先根据空间对象的最大最小窗坐标，检查其是否落在检索范围内：非则排除，是则进一步判断。最后经有关运算，检索出真正落在检索范围内的索引空间对象，并区分完全落入和部分落入两种情况，如图 2-12 所示。

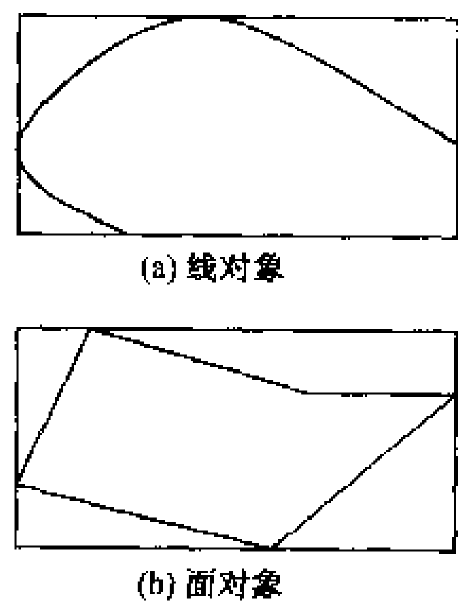


图 2-11　空间对象的窗

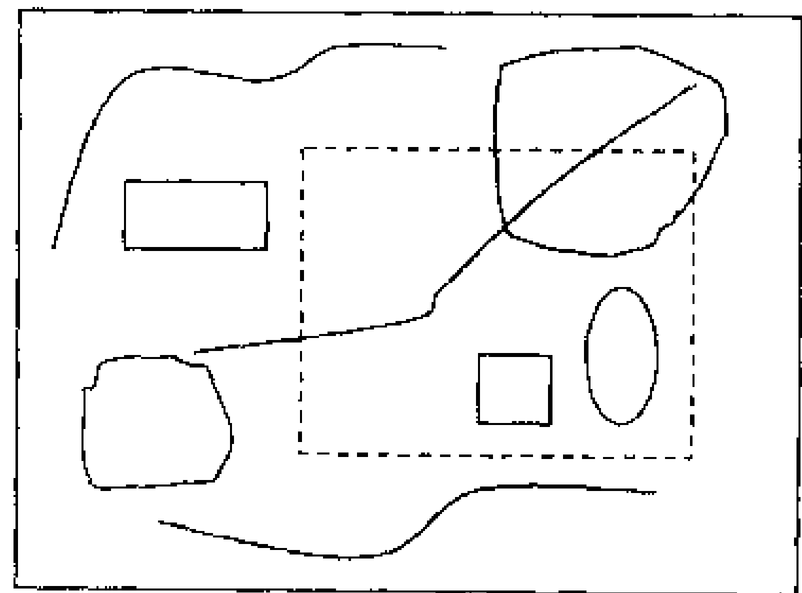


图 2-12　基于窗坐标的空间对象检索

这一方法并没有建立真正的空间索引文件，而是在空间对象的数据文件中增加了其最大最小窗坐标，检索过程还主要依靠空间计算来实现。该机制虽然还需对整个工作区中的空间对象进行窗坐标判断，但毕竟来得更为快捷。而且，随着计算机速度的越来越快，该索引机制还是能满足空间查询的效率要求的。

2. 格网索引机制

格网索引的机制是：事先将工作区按一定的规则划分成格网，并记录每一格网内所包含的空间对象。为了建立空间索引的线性表，可以将格网按 Morton 码进行编码，并建立 Morton 码与空间对象的关系，如图 2-13 所示。

没有包含空间对象的格网为空格网，在索引表中不记录；包含有多个对象的格网，则要分别记录这些对象的标识号，如表 2-5 和表 2-6 所示。如果需要表格化，则要使用串行指针将多个空间目标联系到一个格网内。

21	23	29	31	53	55	61	63
20	22	28	30	52	54	60	62
17	• A 19	25	27	C 49	51	57	59
16	18	24	26	48	50	56	58
5	7	13	15	37	39	45	47
4	6	12	D 14	36	38	44	46
1	3	9	11	33	B 35	41	43
0	2	8	10	32	34	40	42

图 2-13 格网索引编码

表 2-5 对象空间索引

Moton 码	空间对象
8	B
9	B
11	B
12	D
13	D
14	D
15	C, D
19	A
24	C
25	C
26	C, D
27	C
28	C
30	C
33	B
35	B
37	C
38	B
41	B
44	B
48	C
49	C
50	C
51	C
52	C
53	C
54	C
55	C
57	C
60	C
61	C

表 2-6 对象空间索引

空间对象	Moton 码集
A	19-19
B	8-9
B	11-11
B	33-33
B	35-35
B	38-38
B	41-41
B	44-44
C	24-27
C	28-28
C	30-30
C	48-55
C	57-57
C	60-61
D	12-15
D	26-26

3. BSP 树索引机制

BSP 树实质是一种二叉树，它将目标空间逐级进行一分为二的划分，如图 2-14 所示。BSP 树的优点是能较好地与空间数据库中空间对象的分布情况相适应，缺点是其深度较大，不利于空间操作。

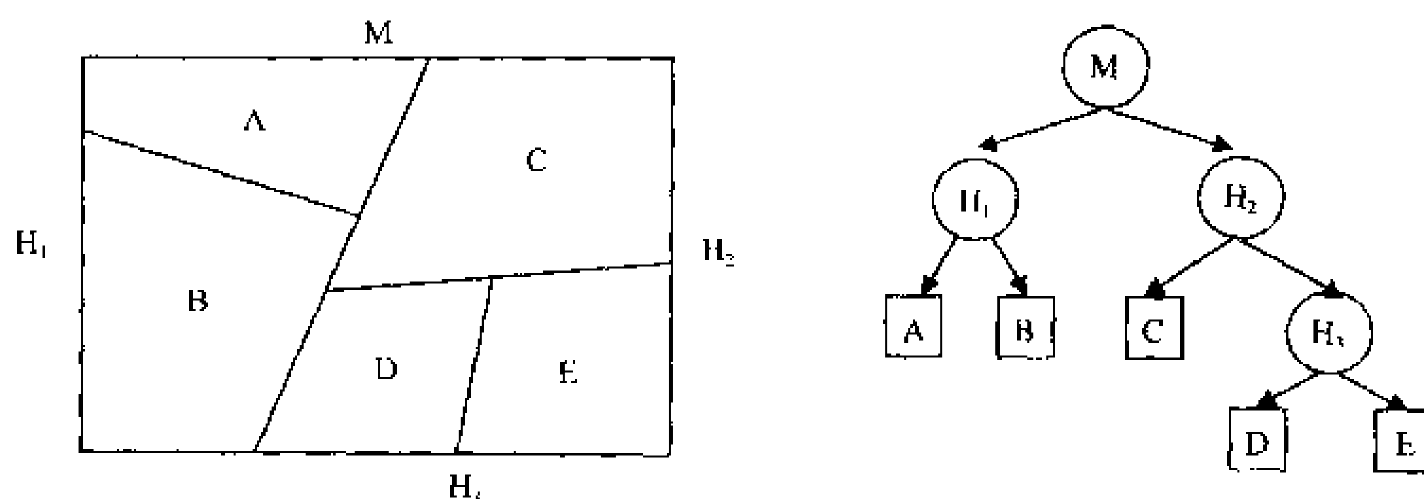


图 2-14 BSP 树索引示例

4. KDB 树索引机制

KDB 树是 BSP 树向多维空间的扩展。KDB 索引的优点是可以对多维空间中的点进行动态索引，删除和增加空间点对象很方便；缺点是不直接支持占据一定时间范围的空间对象，如二维空间中的线与面。针对这一缺点，可以通过空间映射或变换，利用点索引结构来对区域进行索引，进而将原始空间中的区域查询转化为高维空间中的点查询。

5. R 树和 R^+ 树索引机制

R 树索引的机制是：设计一些虚拟的矩形框，将一些空间距离相近的空间对象包含在这些矩形框内。矩形框包含空间对象的指针，因而可以作为空间索引。矩形框的数据结构为：

Rect (Rectangle-ID, Type, Min-X, Min-Y, Max-X, Max-Y)

其中：

Rectangle-ID 为矩形框的对象标识；

Type 表示该矩形框是虚拟空间对象还是实际空间对象；

Min-X, Min-Y, Max-X, Max-Y 为该矩形的最大最小坐标。

构造矩形的原则是：①矩形之间尽可能少重叠；②矩形尽可能包含更多的空间对象；③矩形可以嵌套，即矩形中可以包括更小的矩形。

进行空间检索时，首先判断哪些虚拟矩形落在检索窗口内，然后再进一步判断落在检索窗口内的虚拟矩形中的哪些空间对象为被检索对象。在窗坐标索引方式中，空间对象的窗可以看成是虚拟矩形的极小化，N 个空间对象就对应 N 个虚拟矩形。显然，R 树索引方式相当于增加了一层窗，因而与窗坐标索引方式相比，可以进一步提高检索速度。

但事实上，空间对象千姿百态，其最小矩形窗经常重叠，要使每一个空间对象仅被

一个虚拟矩形所包括，就保证不了虚拟矩形不重叠。为此，国外学者提出了改进的 R 树索引法，称为 R^+ 树索引。

R^+ 树索引的机制是：允许虚拟矩形相互重叠，允许一个空间对象同时为多个虚拟矩形所包含，如图 2-15 所示。

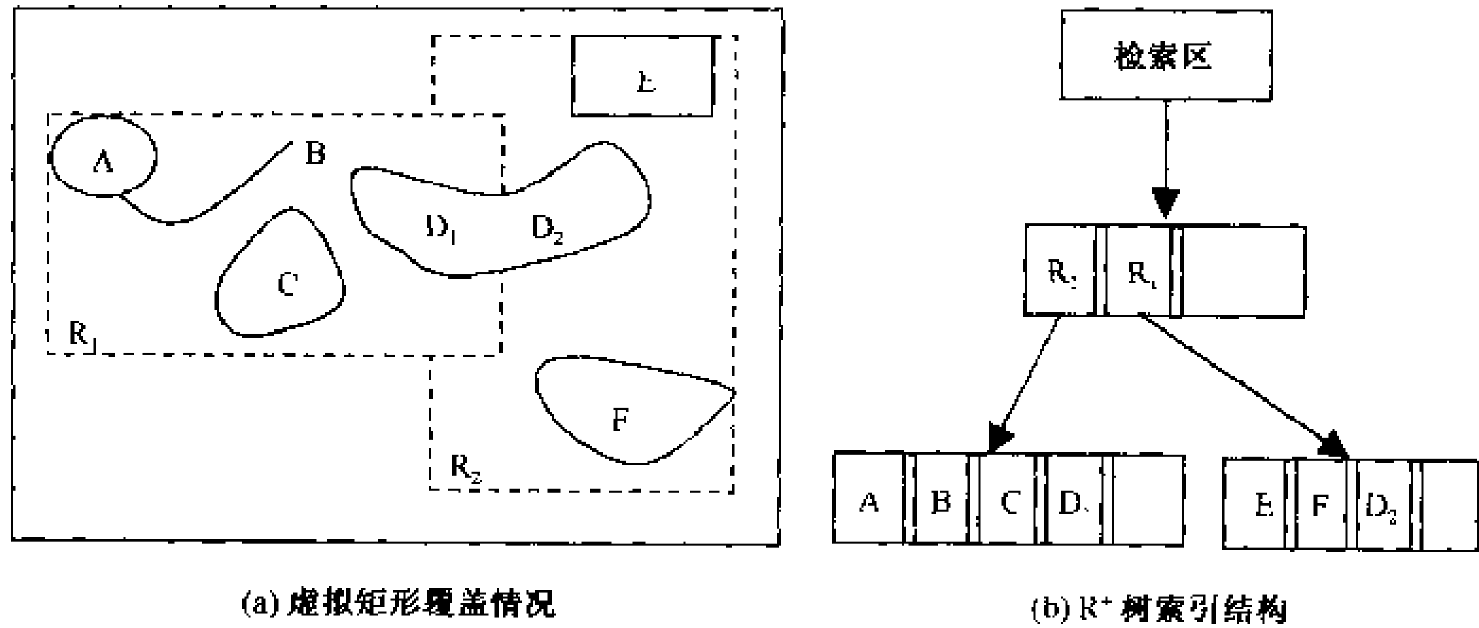


图 2-15　R^+ 树索引实例

尽管 R 树和 R^+ 树索引与窗坐标索引方式相比，检索速度得到提高，但在空间数据的插入、删除和空间搜索效率两方面仍然难于兼顾。

6．CELL 树索引机制

鉴于 R 树和 R^+ 树索引的上述缺陷，CELL 树应运而生。CELL 树索引机制是：借鉴 BSP 树和 R 树机制，采用凸多边形代替矩形作为区域划分的基本单元，子空间不允许重叠，如图 2-16 所示。CELL 树的优点是可以根据图幅范围内的图素分布与复杂情形来自由选择凸多边，因而在某些特殊条件下，CELL 树的磁盘访问次数比 R 树和 R^+

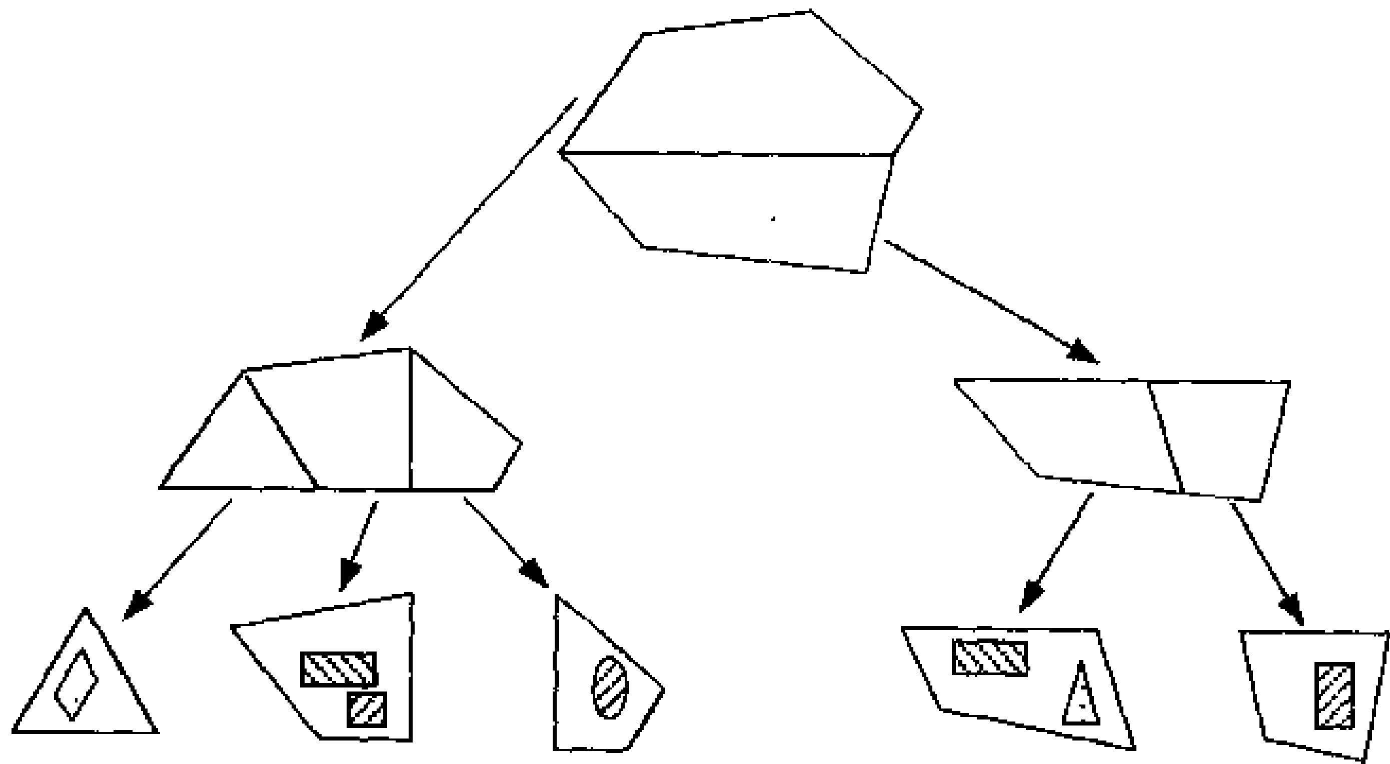


图 2-16　CELL 树索引实例

（据邬伦，2001）

树少。所以，在某些特定条件下，CELL树是一种较优秀的空间索引方法。

参 考 文 献

边馥苓．1996．地理信息系统原理和方法．北京：测绘出版社．118－130

陈述彭，鲁学军，周成虎．1999．地理信息系统导论．北京：科学出版社

龚健雅．2001．地理信息系统基础．北京：科学出版社．191－224

龚健雅．1993．整体SIS数据组织与处理方法．武汉：武汉测绘科技大学出版社

郭达志等．2002．地理信息系统原理与应用．徐州：中国矿业大学出版社

贺建忠．1995．面向对象的GIS空间数据模型的研究（博士学位论文）．北京大学

黄杏元，汤 勤．1989．地理信息系统概论．北京：高等教育出版社

李德仁，龚健雅，边馥苓．1993．地理信息系统导论．北京：测绘出版社

王家耀．2001．空间信息系统原理．北京：科学出版社．108－135

邬 伦等．2001．地理信息系统——原理、方法与应用．北京：科学出版社．44－46，156－158

吴信才．2002．地理信息系统原理与方法．北京：电子工业出版社

张 超，陈丙咸，邬 伦．1995．地理信息系统．北京：高等教育出版社

Atsuyuki O., Barry B., Kokichi S. and Sung N. C., 2000. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, INC

Martien M., 1998. An Introduction to The Theory of Spatial Object Modelling for GIS. Talylor & Francis Ltd

Michael N. D., 2000. Fundamentals of Geographical Information Systems. John Wiley & Sons, INC

Paul A. L., Michael F. G., David J. M. and David W. R., 1999. Geographical Information Systems (I): Principles And Technical. John Willey & Sons, INC

Paul A. L., Michael F. G., David J. M. and David W. R., 1999. Geographical Information Systems (II): Management Issues And Applications. John Willey & Sons, INC

Peter A. B., 1986. Principles of GIS for Land Resources Assessment. Oxford: Clarendon Press

Peter A. B. and Rachael A. M., 1998. Principles of Geographical Information Systems. Oxford University Press

第三章　2D GIS空间数据结构

地学空间对象具有三大特征，即空间特征、属性特征和时间特征。空间特征又分为空间位置和拓扑关系，时间特征表现为空间特征、属性特征的动态变化规律。空间对象的计算机表达即是先用合适的数据模型描述空间对象的关系，然后用合适的数据结构表达空间内对象的三大特征，其中空间对象的形态特征可以由位置特征导出。因此，GIS中主要表达的是空间对象的位置信息、拓扑关系和属性信息。由于空间拓扑关系异常复杂，有些可以通过信息结构进行显式记录；有些则不可以，必须通过空间操作来完成。所以，空间信息表达概括起来包括两个方面内容：其一是记录信息的结构；其二是信息操作的机制。但是，到底哪些空间关系需要预先记录，并存储在数据库中；哪些要通过即时计算，用空间操作来完成，这要由空间关系对于某一领域的重要性，以及操作的复杂程度来决定。随着计算机处理速度越来越快，许多GIS如MapInfo，MGE，ArcGIS等，都把空间关系的表达交给即时计算，而不再主要在数据库中存储。

按照空间对象信息在计算机中的表达方式进行区分，有三种主要数据结构，即：基于坐标的矢量（vector）数据结构，基于规则格网的栅格（raster）数据结构，以及基于铺盖的镶嵌（tesselation）数据结构。除此之外，考虑到矢量数据和栅格数据两者特点的综合，人们还提出了基于矢量和栅格结合的数据结构；以及面向多媒体数据管理的超图数据结构等。

§3.1　矢量数据结构

现实世界中，面域的边界可以用线来描述，而线又可以用坐标串来描述，坐标串即是一些有序的坐标点的集合，这就是矢量描述的基本形式。采用矢量方式描述空间对象时，可以包含拓扑关系，也可以不包含拓扑关系，分别称为无拓扑的矢量数据结构和含有拓扑的矢量数据结构。

3.1.1　无拓扑的矢量数据结构

用无拓扑的矢量数据结构表达的矢量数据模型仅记录空间对象的位置坐标和属性信息，而不记录其拓扑关系，面条数据模型（spaghetti model）是其中的一种。可以适用无拓扑的矢量数据结构的空间对象包括：等高线、等值线、等势线等各类抽象的线形目标。它有两种模式：一种方式是用点、线、面对象分别记录其坐标对；另一种方式是用一个文件记录点坐标对（称为坐标文件），而线、面由点号组成。

若按第一种方式，每个空间对象的坐标均独立存储，不顾及相邻的面、线和点状对象。如此将导致除边界线以外的所有公共边均要存储两次，所有公共点均要存储两次以上。除引起不必要的数据冗余之外，还必然产生数据裂缝、数据重叠和点位不重合，如

图 3-1 所示。为消除这些现象，需要做大量的编辑、检查工作。

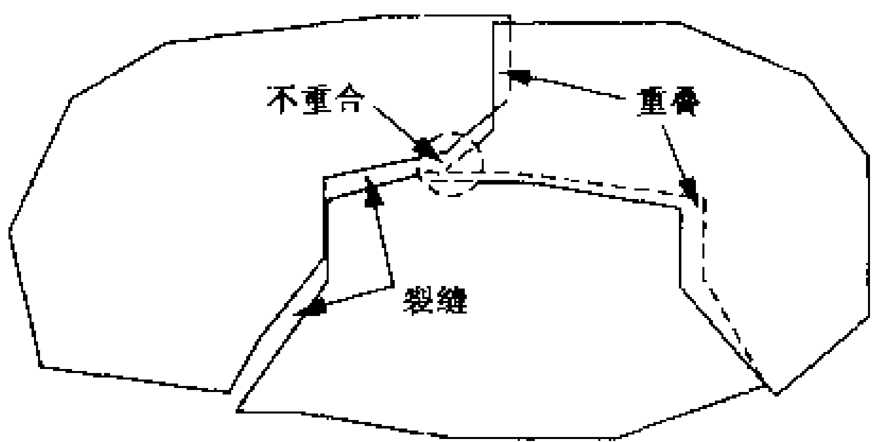


图 3-1 矢量数据的裂缝、重叠与不重合现象

若按第二种方法，由于所有的点号及其点位坐标均在坐标数据文件中记录且仅记录一次，而线、面对象仅记录组成它的点号序列。因此，既避免了数据冗余，也不会引起数据裂缝和重叠，更没有点位不重合的可能。

图 3-2 是两种方式的比较。表 3-1 是对应第一种方式的数据记录；表 3-2 和表 3-3 是对应第二种方式的数据记录。

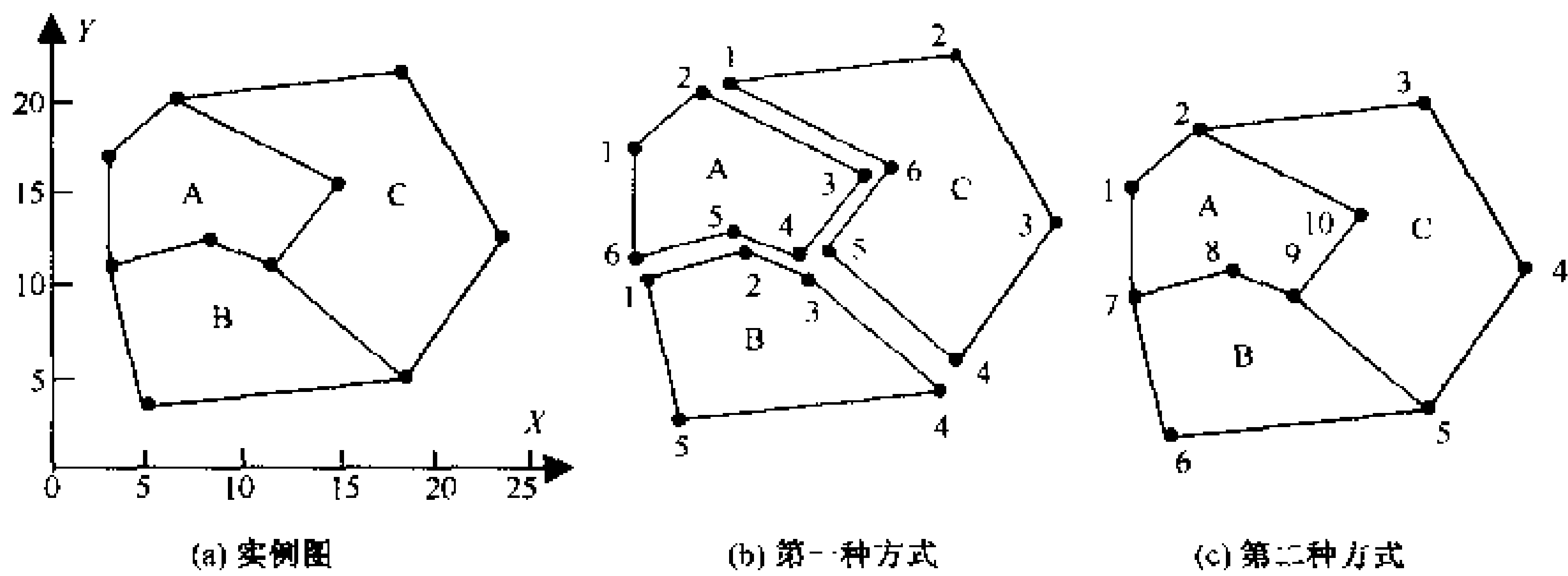


图 3-2 无拓扑矢量数据的两种存储方式比较

表 3-1 多边形数据文件

多边形 ID	多边形用户编码	边界坐标对（x，y）	
A	P101	3.0 7.0 …	17.0 20.0 …
B	P102	3.5 8.0 …	10.0 11.5 …
C	P103	7.5 19.0 …	20.5 22.0 …

表 3-2　点位坐标文件

点　号	坐标对 (x, y)	
1	3.0	17.0
2	7.0	21.5
3	18.5	20.5
…	…	…

表 3-3　多边形文件

多边形 ID	多边形用户编码	边界点串
A	P101	1, 2, 10, 9, 8, 7, 1
B	P102	7, 8, 9, 5, 6, 7
C	P103	2, 3, 4, 5, 9, 10, 2

3.1.2　有拓扑的矢量数据结构

空间拓扑关系是GIS的重要标志（见第五章）。但实际上，目前大部分GIS所存储的拓扑关系仅仅涉及空间对象的拓扑关联关系（如Arc/Info, System 9, DIME和TIGER文件等)。其他拓扑关系如拓扑邻接、拓扑包含可以从关联关系中导出，或通过实时空间运算得到。拓扑关联关系可以有两种表达方式，即全显式表达和半隐含式表达。

1. 全显式表达

全显式表达是指结点、弧段、面域之间的所有拓扑关联关系都用关系表表达出来。如图3-3所示区域，其从上到下（由图区、面域、弧段到结点）的组成关系见图3-4。其结点-弧段-面域的拓扑关联关系可以用表3-4至表3-7完全描述出来。

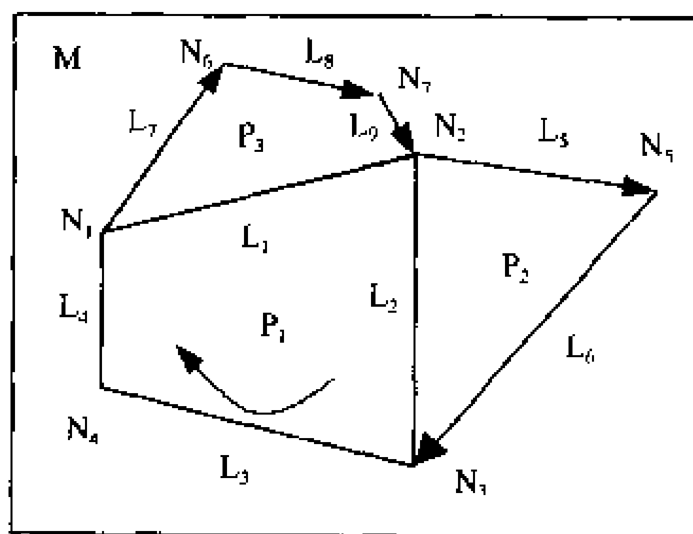


图 3-3　矢量地图实例

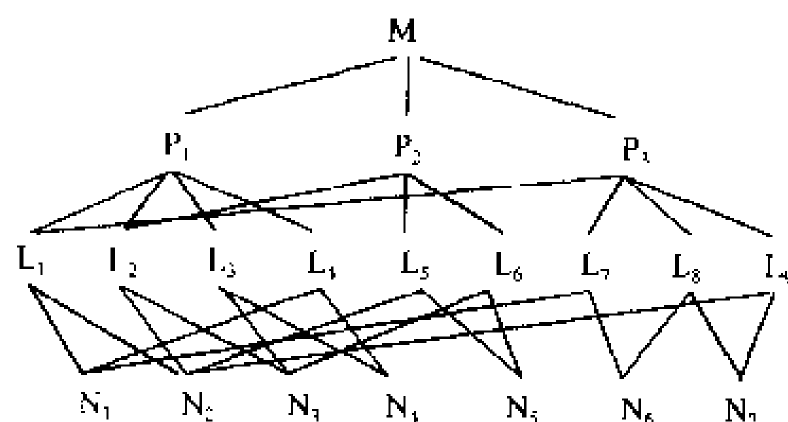


图 3-4　结点-弧段-面域的拓扑关联关系

其中，前两个表描述了从上到下的拓扑关联，后两个表则描述了从下到上的拓扑关联。由表可见，这4个表并没有包括点与面、面与点的直接关联，这种关系是以边线为

桥梁间接建立的。由于表3-5和表3-7均与弧段有关，ARC/INFO通常将这两个表合并在一起，形成表3-8的形式。

表3-4 面域-弧段的拓扑关联关系 $p=p(l)$

面域	弧段
P_1	L_1，L_2，L_3，L_4
P_2	L_5，L_6，L_2
P_3	L_7，L_8，L_9，L_1

表3-5 弧段-结点的拓扑关联关系 $l=l(n)$

弧段	起结点	终结点
L_1	N_1	N_2
L_2	N_2	N_3
L_3	N_3	N_4
L_4	N_4	N_1
L_5	N_2	N_5
L_6	N_5	N_3
L_7	N_1	N_6
L_8	N_6	N_7
L_9	N_7	N_2

表3-6 结点-弧段的拓扑关联关系 $n=n(l)$

结点	弧段
N_1	L_1，L_4，L_7
N_2	L_1，L_2，L_5，L_9
N_3	L_2，L_3，L_6
N_4	L_3，L_4
N_5	L_5，L_6
N_6	L_7，L_8
N_7	L_8，L_9

2. 半隐含表达

如果仅用前面的部分表格来表达空间对象的拓扑关联关系，则称之为半隐含表达（龚健雅，1993）。例如System 9仅用表3-4和表3-5表达从上到下的拓扑关联关系，其他关系由这两个表隐含表达，需要时再建立临时的关系表；而美国人口调查局早期的

DIME 文件仅采用了表 3-8，其他关系也是在需要时临时使用某种运算方法进行查找。

表 3-7　弧段-面域的拓扑关联关系 $l = l\ (p)$

弧　段	左面域	右面域
L_1	P_3	P_1
L_2	P_2	P_1
L_3	—	P_1
L_4	—	P_1
L_5	—	P_2
L_6	—	P_2
L_7	—	P_3
L_8	—	P_3
L_9	—	P_3

表 3-8　弧段-结点-面域的拓扑关联关系 $l = l\ (n,\ p)$

弧　段	起结点	终结点	左面域	右面域
L_1	N_1	N_2	P_3	P_1
L_2	N_2	N_3	P_2	P_1
L_3	N_3	N_4	—	P_1
L_4	N_4	N_1	—	P_1
L_5	N_2	N_5	—	P_2
L_6	N_5	N_3	—	P_2
L_7	N_1	N_6	—	P_3
L_8	N_6	N_7	—	P_3
L_9	N_7	N_2	—	P_3

实际上，不同的 GIS 系统在描述拓扑关联关系时，以上几张表的结构可能略有不同。比如，表 3-4 描述面域与弧段的关系、表 3-6 描述结点与弧段的关系时，必然出现变长记录。对于该问题的处理，美国计算机图形及空间分析实验室的办法是采用串行指针的办法，而 Arc/Info、Geostar 等软件则采用直接存储变长记录的办法。

有时候，根据专题 GIS 的特殊需要，可以有选择地建立拓扑关系来简化系统的数据准备工作。例如，对于一个主要以面域对象为主体的国土资源管理信息系统，表 3-6 所表达的拓扑关系可以省略；但对于一个交通 GIS，表 3-6 则不可或缺。但无论如何，表 3-4 所表达的拓扑关系必不可少，否则谈不上建立拓扑关系，因为这是建立多边形公共边数据共享的桥梁。

§3.2　栅格数据结构

除了用点的显式坐标方式来记录和表达现实世界之外，还可以用另外一种隐含坐标

的非显式方式来记录和表达现实世界，即将连续区域离散化，用二维规则栅格覆盖整个连续区域。现实生活中，有很多这样的例子，如我们所熟悉的照相底片、冲洗出来的照片、报纸和书刊印刷出来的图片、卫星遥感影像、航空摄影照片以及扫描仪扫录的结果数据等等。

3.2.1 栅格像元的取舍法则

实际上，栅格数据就是像元（pixel）阵列。每个像元由行列号惟一地确定其所在平面位置，给像元赋予属性以表达该像元所覆盖的空间实体的类型或属性值的编码。点实体在栅格数据结构中惟一由一个像元表示；线实体则由一系列相互连接的像元串的集合组成；面像元则由聚集在一起的相邻像元团块表示。如图 3-5 所示。

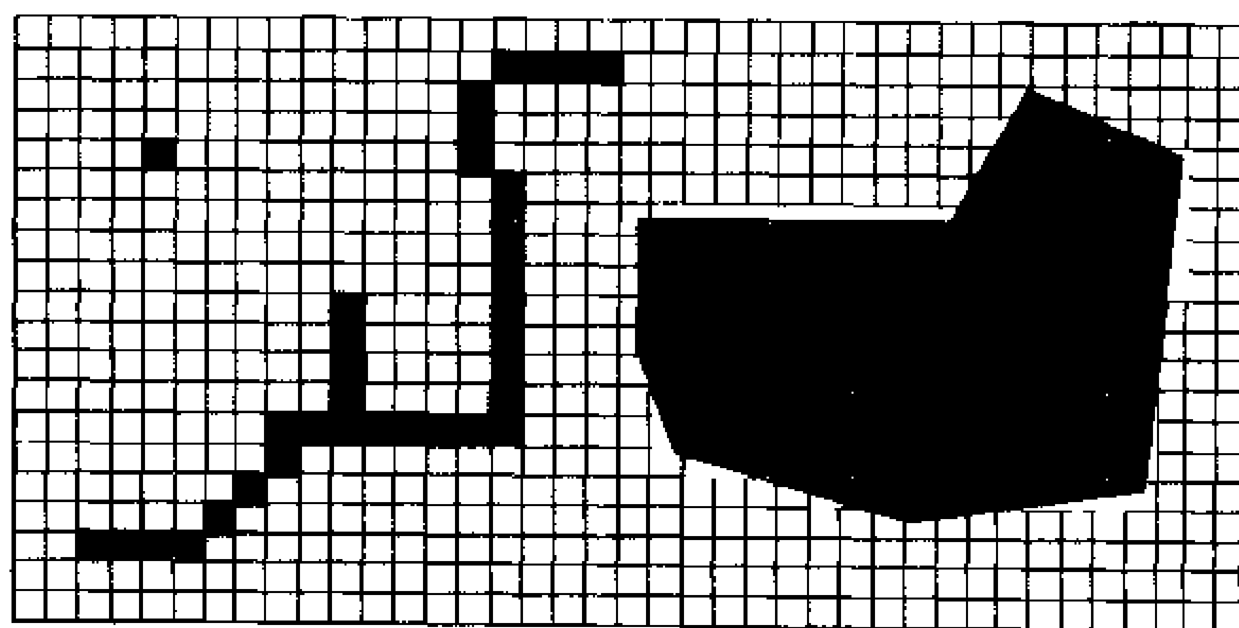


图 3-5　点、线、面实体的栅格数据结构表达

面域的边界不会总是恰好沿行列线行进，必然导致面域的边界穿越某些像元。在像元取舍时，一般遵循面积占优法则：即若该像元中被某面域所覆盖的面积超过一半，则该像元为面域像元；否则，为非面域像元。如图 3-5 所示，沿面域边界的灰色像元即为根据面积占优法则选择的有效像元。

弧线的边界也并不总是与行列方向平行，此时像元取舍时应遵循以下原则：

(1) 后续像元应与前面已确定的像元边相连或角相连；

(2) 沿非主方向上不能出现边邻接像元。

此处所谓主方向，是指弧线局部位置的优势方向。若弧线局部切线与行线的夹角小于 45°，则该弧线的优势方向为行线方向；若弧线局部切线与列线的夹角小于 45°，则该弧线的优势方向为列线方向。

如果某条弧或面域的某一边界线的某一局部（或其整体）与行（或列）线平行，并且恰好与行（或列）线重合，此时决定是选择该行线的上侧像元还是下侧像元（或该列线的左侧像元还是右侧像元），应遵循非凸法则：即不形成孤立的凸起像元组。如图3-6所示，其中灰色像元为选择像元，斜线像元为舍弃像元。

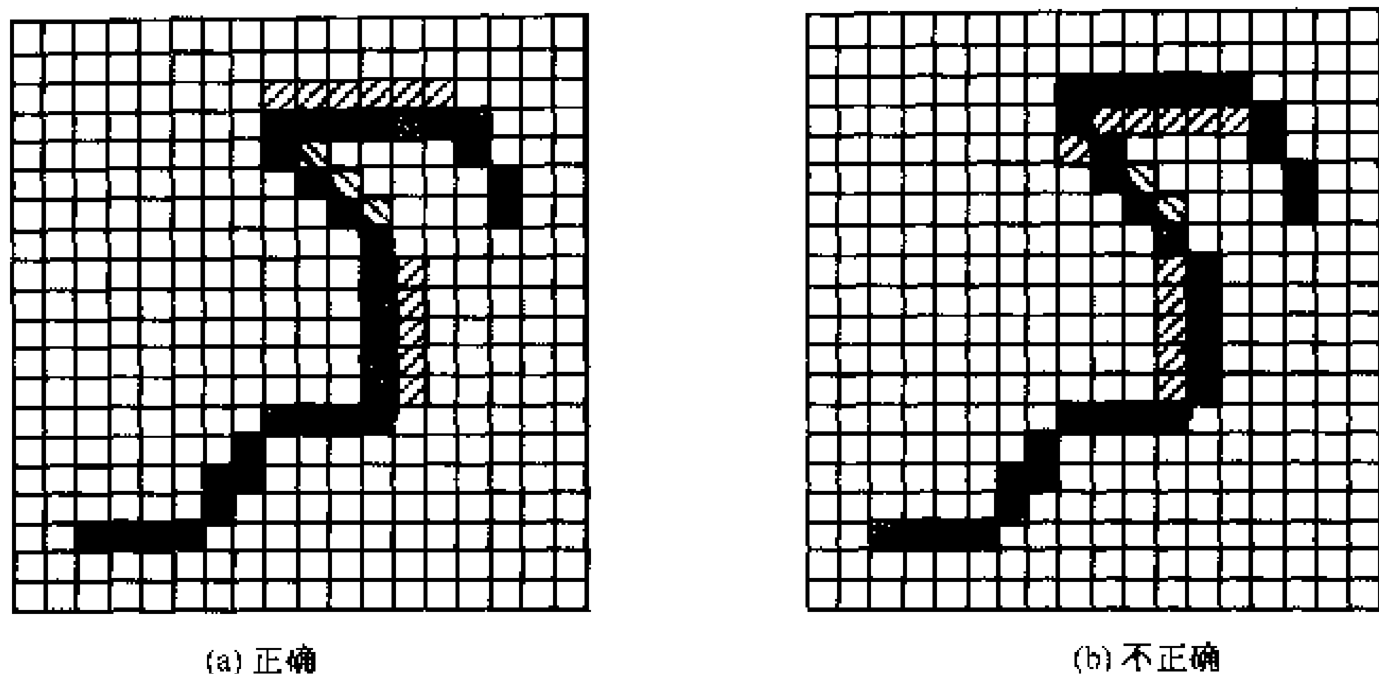


图 3-6 线实体栅格像元的选择

3.2.2 栅格尺寸的选择

由图 3-5 和图 3-6 可以看出，空间对象的几何位置和形态表达的精确性完全取决于栅格的尺寸。栅格尺寸越大，精确性越差，甚至某些小面积的空间对象将反映不出来；反之，则精确性越好，线和面域轮廓越接近于矢量描述结果。但是，随着栅格尺寸的缩小，单位面积内的栅格像元数量将几何级数增加，必然带来数据文件的膨胀。即使采用某种压缩存储算法（参见第十章），也无法扭转数据膨胀的趋势。

以桌面扫面仪进行黑白图像扫描为例，现有激光打印机输出的“MGIS”四个英文字母，扫描范围宽约 1.5cm。分别按不同扫描分辨率扫描后得到的栅格图像及其压缩文

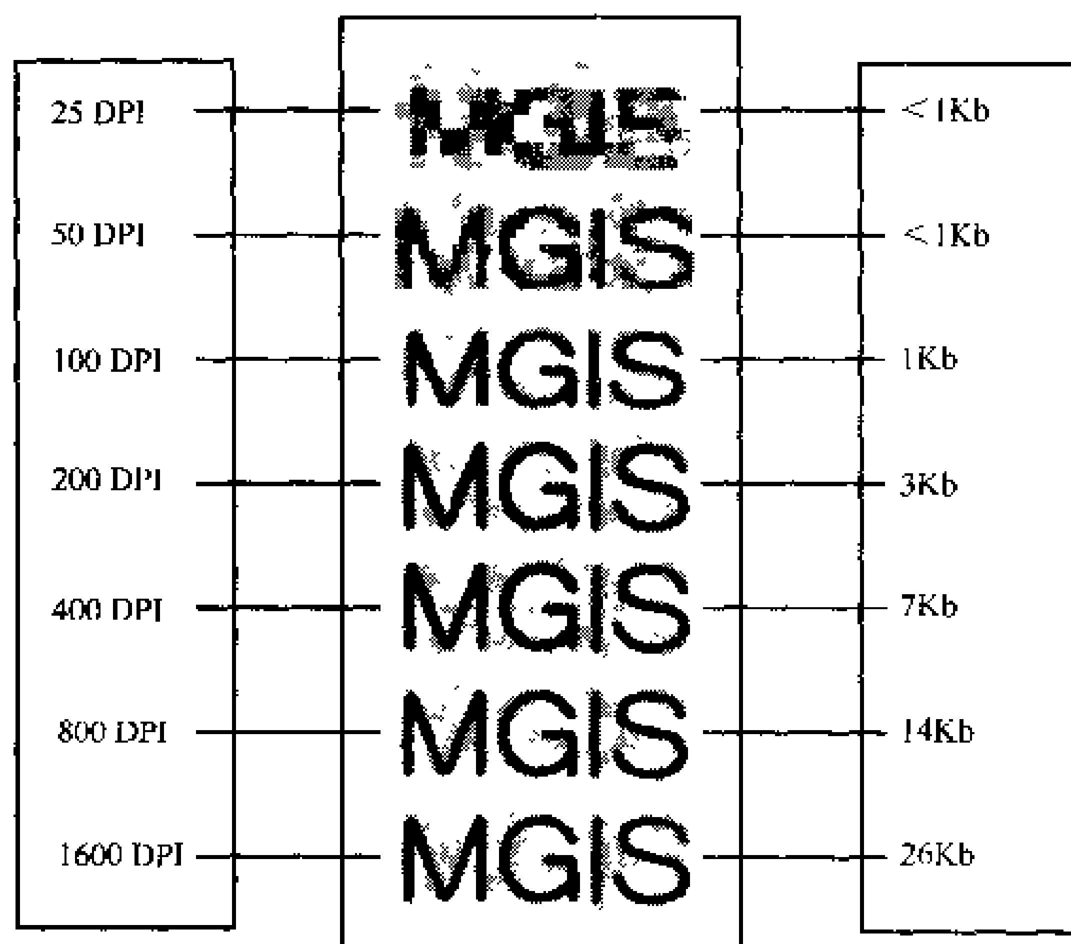


图 3-7 栅格像元尺寸与表达精度、文件大小的关系实例

件（JPEG 格式）大小见如图 3-7 所示。

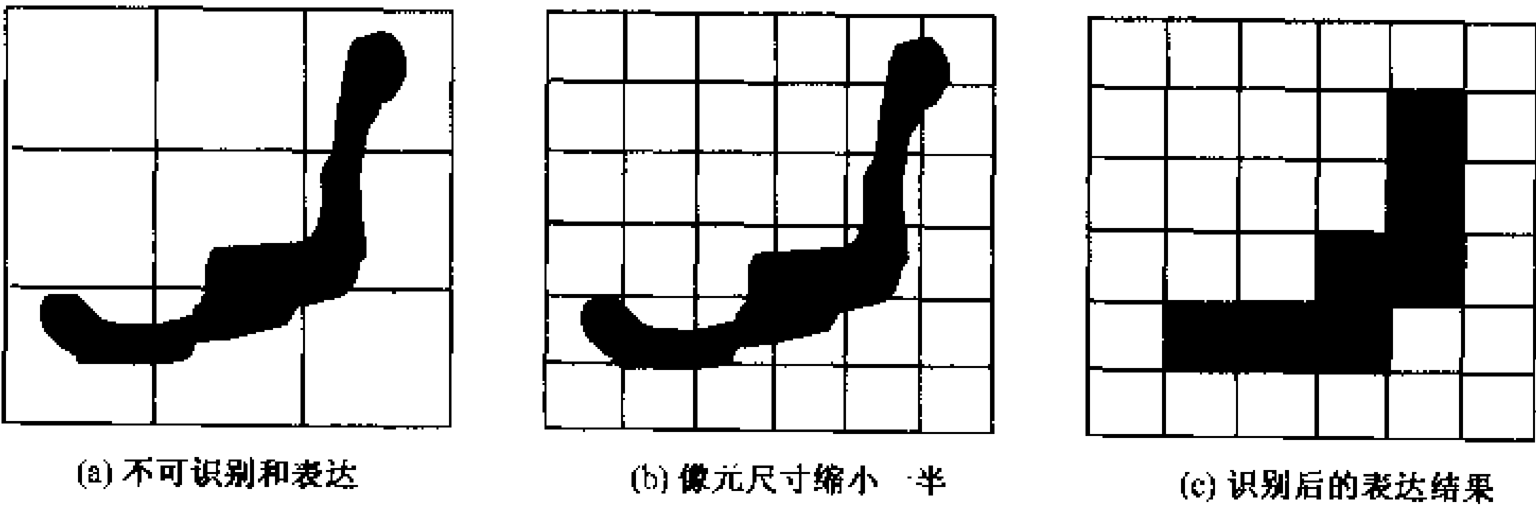
再以 1 km^2 范围的飞星遥感影像数据为例，假设每个像元需要 8 个字节（byte）来记录其位置编码和遥感信号，如果忽略因空间分辨率的提高所引起的遥感信息内容的变化，则该范围内的遥感影像文件的大小理论上（记录时无压缩）将与遥感空间分辨率直接相关，如表 3-9 所示。

表 3-9　遥感影像文件大小与空间分辨率对比

空间分辨率	影像文件大小
10 m	80 K
1 m	8 M
0.1 m	800 M
0.01 m	80 G

理论上来说，为了最大限度地逼近空间对象的分布形态，栅格像元的尺寸应越小越好。但这样做必然使数据量激增，并产生大量的冗余数据。为此，如何选择合适的栅格尺寸，不仅对于空间对象识别，而且对于空间数据文件存储均至关重要。

通常，根据研究区域内所要识别的最小面状目标的面积来确定栅格的像元尺寸。如图 3-8 所示。



(a) 不可识别和表达　(b) 像元尺寸缩小一半　(c) 识别后的表达结果

图 3-8　长岛形对象的栅格表达与像元尺寸的关系

设当前研究区域中，所需识别的最小面状对象（如建筑物、池塘、草坪、井盖等）的面积为 A_{min}，则栅格像元的尺寸 S_r 至少应为：

$$S_r = 0.5\sqrt{A_{min}} \tag{3-1}$$

由于空间对象的边界十分复杂，某些面域对象可能比图 3-8 中的长岛还要狭长。此时，即使按公式（3-1）求出的栅格像元尺寸，仍然不能识别和表达它。那么，就要根据这些特殊地物对象的形态特征及所需识别精度的要求，进一步降低栅格像元的尺寸，比如降低 25%、一半甚至更多。

3.2.3 栅格数据的组织

栅格数据采用一组笛卡尔平面来描述空间对象的属性。所谓笛卡尔平面，实际是一个二维数组，其中数组中的某行某列即对应现实世界中的某一栅格单元的某一项属性值。通常，也称这种平面为“层”。如图 3-9 所示为某一地区的地形地貌及其数据层；图 3-10 分别表示这些数据层的所有笛卡尔平面的叠加。

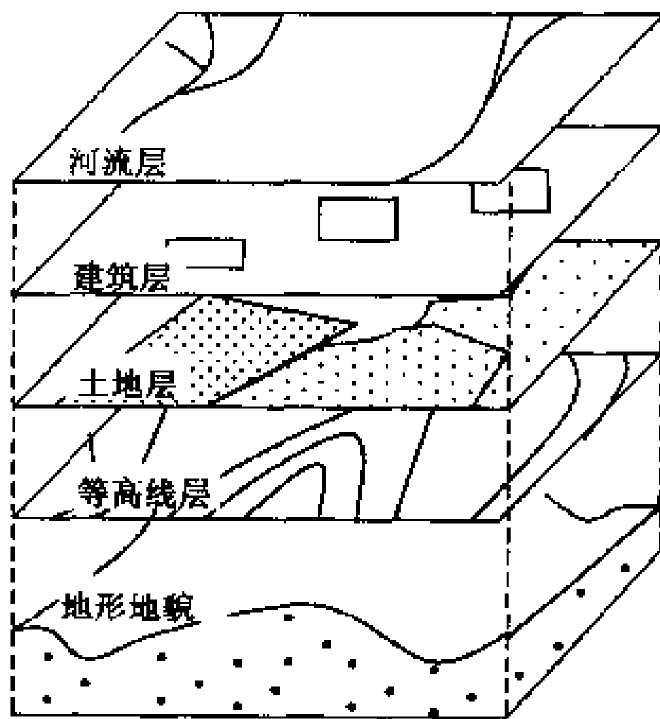


图 3-9 地形地貌及其数据层

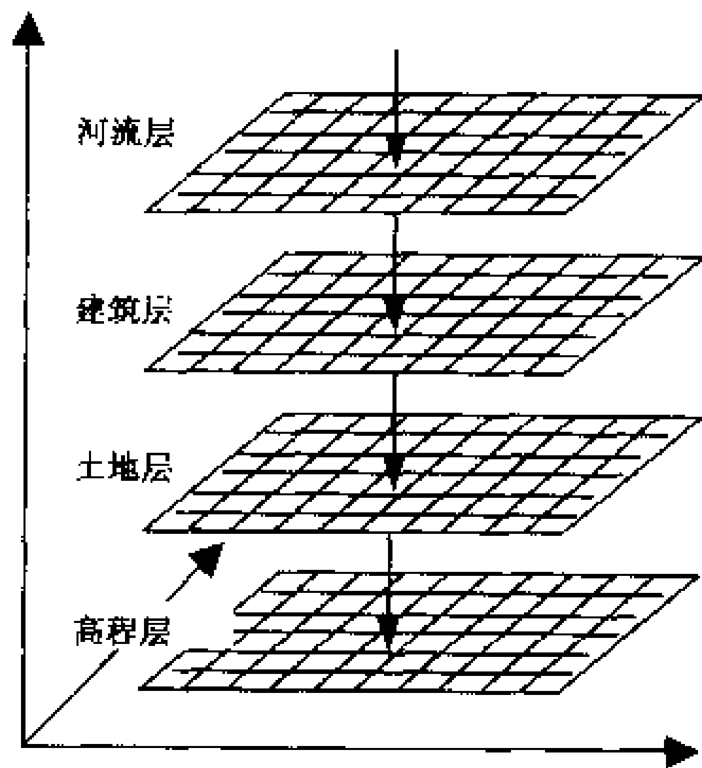


图 3-10 笛卡儿数据平面

1．栅格数据组织方式

在数据库中，栅格数据的组织有三种基本方式：基于像元、基于层和基于面域。

(1) 基于像元：以像元作为独立存储单元，每一个像元对应一条记录，每条记录中的记录内容包括像元坐标及其各类属性值的编码。

(2) 基于层：以层作为存储基础，层中又以像元为序记录其坐标和对应该层的属性值编码。

(3) 基于面域：也以层作为存储基础，层中再以面域为单元进行记录，记录内容包括：面域编号、面域对应该层的属性值编码、面域中所有像元的坐标。

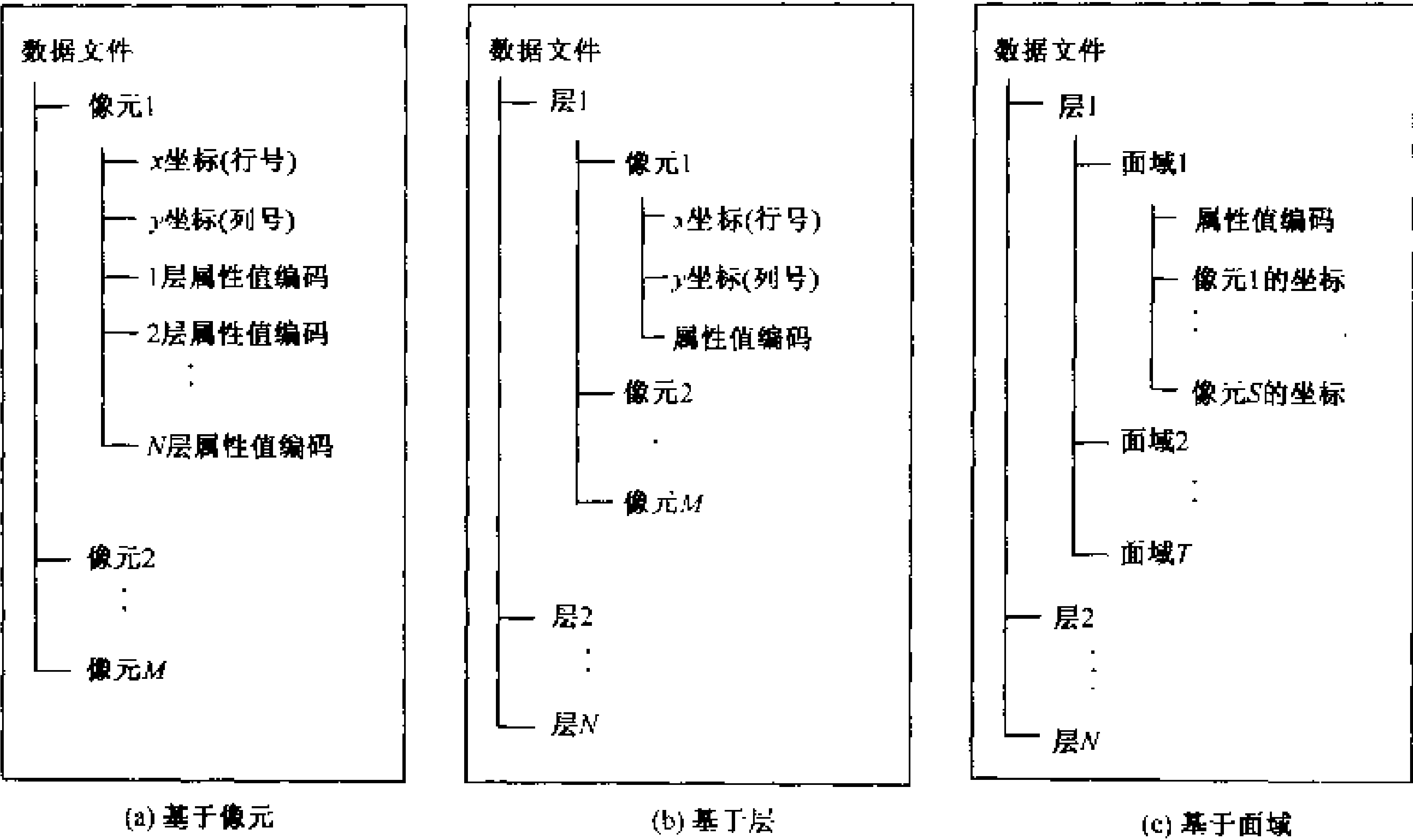
上述三种栅格数据组织方式的对比如图 3-11 所示。

由此可见，基于像元的数据组织方式简单明了，便于数据扩充和修改，但进行属性查询和面域边界提取时速度较慢；基于层的数据组织方式便于进行属性查询，但每个像元的坐标均要重复存储，浪费了存储空间；基于面域的数据组织方式虽然便于面域边界提取，但在不同层中像元的坐标还是要多次存储。

2．栅格数据组织顺序

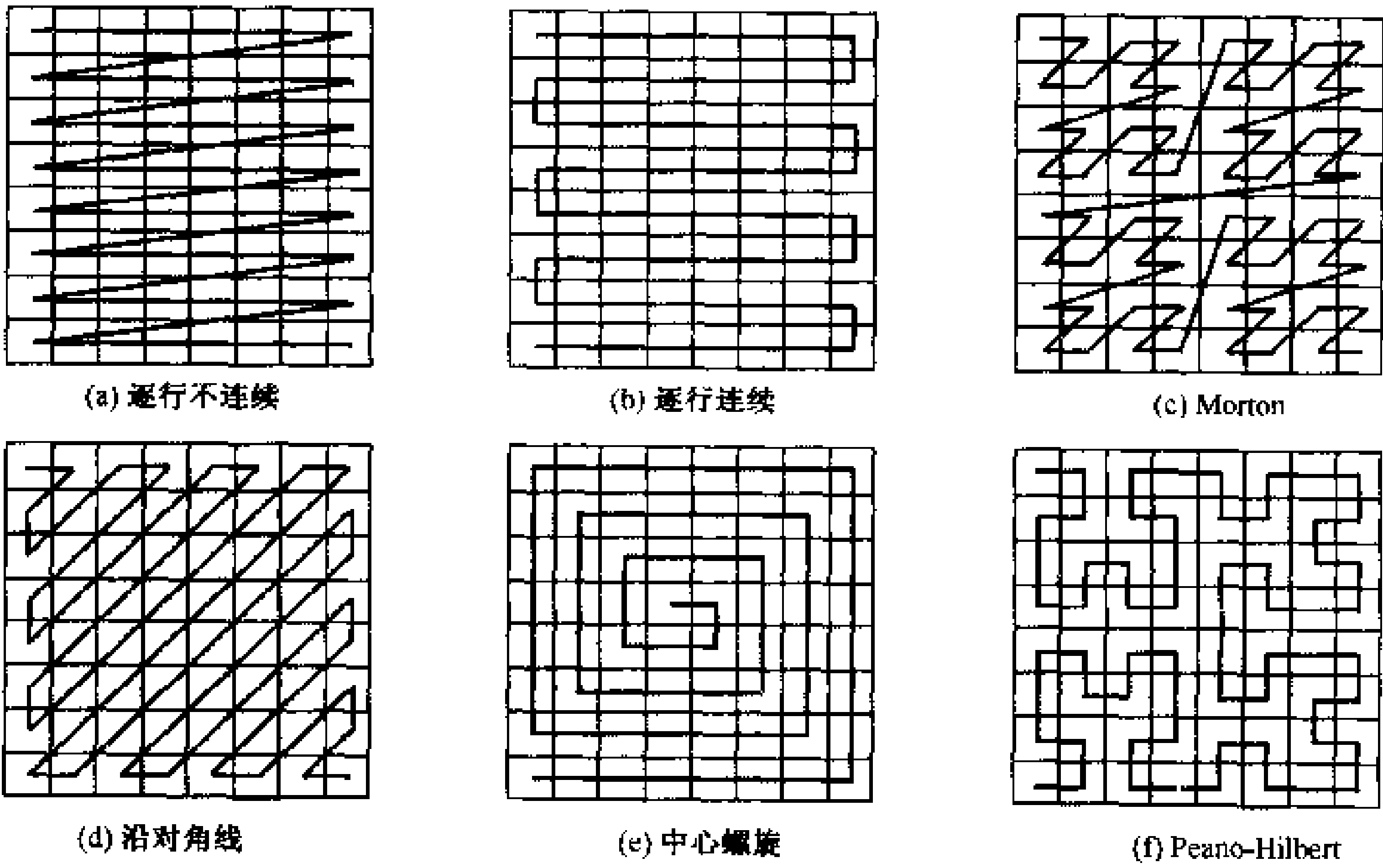
无压缩情况下，栅格数据按直接编码顺序进行组织。所谓直接编码，是指将栅格数据看成一个数据矩阵，逐行（或逐列）逐个记录各栅格的代码，可以是每行均从左到右

逐个记录，也可以是奇数行从左到右而偶数行从右到左逐个记录。有时，针对特定的对象或为了特定目的，可以采用如图 3-12 所示的特殊的记录顺序。



(a) 基于像元　(b) 基于层　(c) 基于面域

图 3-11　栅格数据结构的数据组织方式



(a) 逐行不连续　(b) 逐行连续　(c) Morton

(d) 沿对角线　(e) 中心螺旋　(f) Peano-Hilbert

图 3-12　栅格数据结构的数据组织顺序

§3.3 矢栅结合的数据结构

矢量、栅格数据结构各有其优缺点。矢量数据结构虽然表达地理数据的精度较高，图形输出美观，数据量小，拓扑关系描述完整。但是，数据结构复杂，面域叠加算法较为复杂。而栅格数据结构虽然结构简单，面域叠加和地形表面模拟方便，但数据量大，精度较差，图形输出不够美观，尤其难以进行网络分析，投影变换花的时间也较多。

矢量数据结构的主要优势还在于：矢量结构的基本逻辑单位是空间实体，如建筑、河流、道路等，通过显式地给定空间实体一个惟一的标识符（或逻辑编号），就可以赋予每个空间实体以个性化的属性描述和拓扑关系表达；这不仅便于采用面向对象技术建模，还便于空间查询和空间分析。因此，目前大部分的 GIS 软件均采用矢量数据结构。

尽管如此，栅格数据结构也还是十分重要，尤其在与遥感、数字高程模型（Digital Elevation Model，DEM）的结合方面，具有不可替代的作用，而且这是 GIS 与 RS 集成的重要桥梁。

因此，充分利用两者的优点，在同一个系统中将两者结合起来，是现代 GIS 发展的一个重要的理论与技术议题。目前，主要提出了两种矢栅结合的模式，其一为矢栅混合，其二为矢栅一体化。

1. 矢栅混合模式

矢量栅格混合有多种形式。其中最简单也最实用的是不对矢量结构的数据和栅格结构的数据作任何特殊处理，直接将它们分别存储在同一个 GIS 的空间数据库系统中，并通过共同的 ID 号将各空间对象的矢量数据、栅格数据及属性数据关联在一起，如图 3-13 所示。需要时，只要将两套数据分别调入，然后就可以进行统一显示、查询和分析。例如，遥感影像、航空影像或扫描的栅格地图，均可以作为背景层放在矢量数据层之下，来增强数据的可视效果和进行直观的规划设计。这种结合方式，已成为 GIS 软件的一个必备功能。

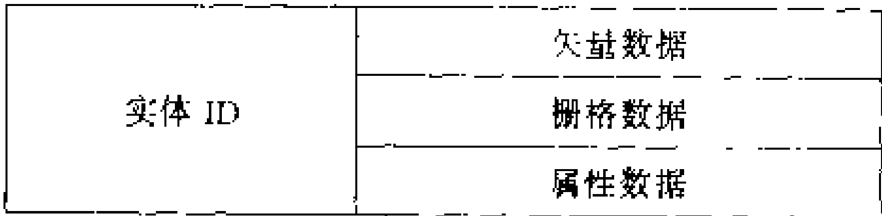


图 3-13 矢栅混合模式的数据组织

矢栅混合模式的缺点是：矢量和栅格两套数据均要无遗漏地在系统中存储，会给系统的存储空间带来压力。

2. 矢栅一体化模式

为了解决矢栅混合增加存储空间这一问题，并更加有效地将矢量、栅格数据结构结合起来，龚健雅（1993）研究提出了矢栅一体化模式。其理论基础是：多级格网方法、三个基本约定和线形四叉树编码。

多级格网方法的实质是将传统的单一格网尺寸划分模式改变为多级格网划分模式。用户可以根据空间对象分布密度、复杂度及地理识别精度要求，将研究区域划分为粗格网、基本格网和细分格网三个层次。其中，粗格网用于建立空间索引；基本格网的大小与常规栅格划分要求一致；细分格网是在点、线经过的基本栅格上再进一步划分为 16 ×16 或 256×256 的小格网，以增加栅格的空间分辨率，从而提高点、线表达的精度。

采用多级格网模式时，粗格网、基本格网和细分格网据均采用线形四叉树编码，并用三个 Morton 码（M_0，M_1，M_2）（参见第十章）表示。其中 M_0 表示点所在或线所通过的粗栅格的 Morton 码，是研究区内的整体编码；M_1 表示点所在或线所通过的基本栅格的 Morton 码，也是研究区内整体编码；M_2 表示点所在或线所通过的细分栅格的 Morton 码，是基本栅格内的局部编码。

以上编码是基于栅格的，因而据此设计的数据结构必定具有栅格的性质。为了使之具有矢量的特点，龚健雅（1993）提出了点状地物、线状地物和面状地物的三个约定：

约定一：点状地物仅有空间位置而无形状和面积，在计算机中仅有一个位置数据；

约定二：线状地物有形状但无面积，在计算机中需要组织一组元子（即栅格单元）填满的路径表达；

约定三：面状地物有形状和面积，在计算机内有一组元子表达的填满路径的边界线和内部（空洞除外均填满，作者注）的区域组成。

据此，点状地物、线状地物和面状地物的“矢量化”数据记录方式如下：

点状地物：如 3.1.2 节中点结构的形式，但以（M_1，M_2）代替（x，y）：

点 ID	M_1	M_2	关联的弧段

线状地物：按 3.1 节中线结构的形式，用 Morton 码代替（x，y）记录原始采样的“拐”点（即中间点）的位置；必要时，还可记录线目标所穿过的所有基本格网的交线位置：

弧 ID	起点 ID	终点 ID	左域 ID	右域 ID	中间点坐标（M_1，M_2）序列……	……

面状地物：除用 Morton 码代替（x，y）记录面状地物边界原始采样的“拐”点（即中间点）位置，以及它们所穿过的所有基本格网的交线位置之外，还要用链指针记录多边形的内部栅格。必要时，还可以记录边界所穿过的所有基本格网的交线位置：

面域 ID	边界 ID 序列	面域内点指针	……

面域内点指针位置	面域内点坐标（M_1，M_2）序列……

因此，点状地物、线状地物和面状地物不仅具有如同矢量数据结构的位置“坐标”，而且还可以有类型编码、属性值和拓扑关系，因而具有完全的矢量特性。与此同时，由于用栅格元子表达了点、填充了线形目标、多边形边界及其内部（空洞除外），实际是进行了栅格化，因而可以进行各种栅格操作。

比较起来，应该说矢栅一体化模式是一种有前途的矢栅结合方法，值得进一步研究和软件实用化。

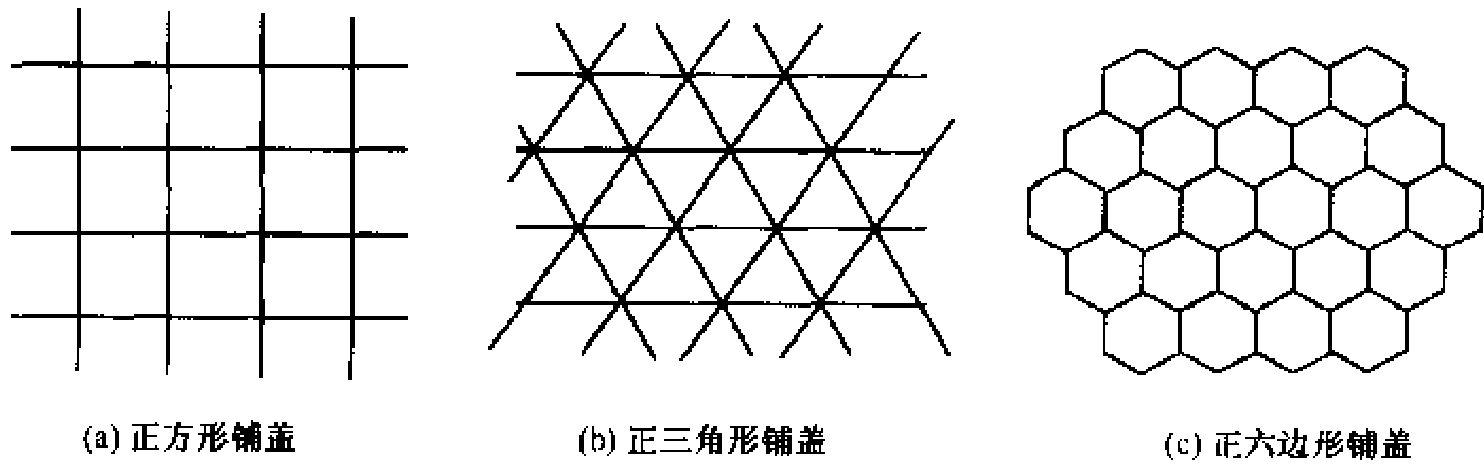
§3.4 镶嵌数据结构

镶嵌（Tessellation）数据结构是基于连续铺盖的，即用二维铺盖或划分来覆盖整个区域。镶嵌是矢量结构的逻辑对偶，有时也称为多边形网格模型（polygon mesh model）（Donna J.P.，1984）。铺盖的特征参数包括尺寸、形状、方位和间距。对同一现象可以有若干不同尺度、不同聚分性（Aggregation or Subdivision）的铺盖。镶嵌数据结构包括规则镶嵌数据结构和不规则镶嵌数据结构，特别适应于二维离散点状空间数据的表达。规则镶嵌最典型的应用模型是格网数字高程模型，其中基于正方形铺盖的栅格数据结构为规则铺盖的特例；不规则镶嵌最典型的数据结构是 Voronoi 图和 Delaunay 不规则三角网，可以当作拓扑多边形处理。

镶嵌是一个很活跃的研究领域，近年各国学者围绕镶嵌理论、技术与方法进行了大量研究（Atsuyuki O. et al.，2000），包括地图的矢量分割与栅格分割、2D 镶嵌与 3D 镶嵌等。Lee Y.C.（2000）对空间镶嵌进行了分类研究，提出特征为主（feature primary）和空间为主（space-primary）两种镶嵌单元，其中前者主要是不规则形状的，而后者则可以分为无约束和受约束两类（受约束类镶嵌可以是层次的或非层次的）。

3.4.1 规则镶嵌数据结构

所谓规则镶嵌数据结构，即用规则的小面块集合来逼近自然界不规则的地理单元。在二维空间中虽有多种可能的规则划分方法（如图 3-14 所示），但为了便于有效地寻址，网格单元必须具有简单的形状和平移不变性。图 3-14 所示的 3 类规则铺盖中，只有正方形与正六边形既是规则的又是可平移的，即在整个平面上具有相同的方向。正六边形有 6 个最近的邻域，比只有四个邻域的正方形有更好的邻接性。然而，正六边形的层次性较差，即它不能无限地被分割；而正方形具有无限可分性，是分割二维空间的实用形式，很多环境监测数据的采集和图像处理普遍采用正方形面元（像元）。



(a) 正方形铺盖　　(b) 正三角形铺盖　　(c) 正六边形铺盖

图 3-14　规则的平面铺盖

构造规则镶嵌的具体做法是：用数学手段将一个铺盖网格叠置在所研究的区域上，

把连续的地理空间离散为互不覆盖的面块单元（网格）。划分之后，简单化了空间变化的描述，同时也使得空间关系（如毗邻、方向和距离等）明确，可进行快速的布尔集合运算。在这种结构中每个网格的有关信息都是基本的存储单元。

从数据结构上看，规则网格系统的主要优点在于其数据结构为通常的二维矩阵结构，每个网格单元表示二维空间的一个位置，不管是沿水平方向还是沿垂直方向均能方便地遍历这种结构。处理这种结构的算法很多，并且大多数程序语言中都有矩阵处理功能。此外，以矩阵形式存储的数据具有隐式坐标，不需要进行坐标数字化；规则网格系统还便于实现多要素的叠置分析。因而，规则铺盖是一种重要的空间数据处理工具。

3.4.2 不规则镶嵌数据结构

不规则镶嵌数据结构是指用来进行镶嵌的小面块具有不规则的形状或边界，其典型数据结构是 Voronoi 图和 Delaunay 不规则三角网。

Voronoi 图是俄国数学家 M.G. Voronoi 于 1908 年发现的几何构造，并以他的名字命名。因早在 1850 年，另一位数学家 G.L. Dirichelt 同样研究过这种几何构造，故有时也称其为 Dirichelt 格网。由于 Voronoi 图在空间剖分上的等分性特征，使它在许多领域获得应用，很多几何问题可用 Voronoi 多边形得出有效、精致、在某种程度上还可以说是最佳的解。如果把空间邻接定义为多边形邻接，并把围绕各物体的 Voronoi 多边形的边界用等距离准则来确定，则所有地图上的物体（此处为点和线段）就具有明确的邻居关系。从这一思想出发，就可导出一种统一的途径来处理许多空间问题。Voronoi 图的构建算法详见第十六章。

Delaunay 三角网是俄国数学家 B. Delaunay 于 1934 年发现的。Delaunay 三角网是 Voronoi 图的对偶，即将 Voronoi 图中各多边形单元的内点两两连接后得到一个布满整个区域而又不互相重叠的三角网结构。Delaunary 三角网在地理信息系统或者数字高程模型 TIN 模块中广泛使用，可以从 Voronoi 图中构造，也可以根据离散点直接构造（详见第十五章）。不规则三角网数据结构的核心是三角形的记录。一般采用 3 张表来完成不规则三角网的数据记录：其一记录各三角形的相邻三角形编号，其二记录各三角形的 3 个结点的编号，其三记录各结点的坐标值。

不规则镶嵌的优点为：能依据区域中数据事件的密度来调整格网形状和大小，以适应采样数据的实际分布，进而消除数据冗余。数据越稀，则单元越大；数据越密，则单元越小。不规则镶嵌的缺点为：①生成过程复杂费时；②覆盖操作极为困难。因此，镶嵌网格除了用于一些特定场合以外，作为 GIS 数据库的实用空间数据模型还需要作进一步的研究。

§3.5 超图数据结构

20 世纪 90 年代以来，计算机技术得到飞速发展。人们提出超文本（Hypertext）和超媒体（Hypermedia）等新的结构方式来贴近人的联想式思维方式。在超文本和超媒体模式下，属性数据部分按相互关系组织成正文网，可以从任意一点以联想方式浏览查

阅，其中包括图像、图形、声音、动画和连接播放等多媒体（Multimedia）方式。其中，最基本的组成部分是节点（node）和链（link）。节点用来存储对象的信息，链用来表示信息之间的关系。节点又分为正文节点、图像节点、声音节点、图形节点，链可包括导航链、名称链、聚焦链、全景链、组织推理链、索引链、属性链、组成链、隐含链等，从而构成一个更灵活的查询环境和智能化的系统。超图数据结构（Hypergraph Based Data Structure，HDBS）正是在这一背景下诞生的。

3.5.1 超图数据结构的概念

HDBS最早是由法国巴黎大学的F.Bouille教授于1979年提出来的。后来经过中国学者崔伟宏（1995）和美国的Rugg教授等的研究逐步完善起来。

HDBS是建立在超图和集合论的基础上，基于现象（Phenomenon based）的拓扑数据结构，通常也称为超关系模型（Hyper-Relational Model）。它以"现象结构"理论取代传统的关系结构（Relational Structure）、层次结构（Hierarchical Structure）和网络结构（Network Structure）理论；以类别、目标、属性和关联这4种抽象数据类型作为最基本的结构框架，来描述和表示事物实体及它们的空间与非空间属性。HDBS为3S（RS，GPS，GIS）技术集成和建立一体化数据结构提供了重要的理论基础和有效的可操作的手段。Laurinni早在1992年就预言："在不久的将来，超图将广泛应用于空间信息系统中，或者更确切地说，未来的空间信息系统将建立在超图概念的基础上。"

HDBS通过使用类别、目标、属性和关联这4种抽象数据类型，为表示基本地理实体、复合地理实体以及它们的空间与非空间属性提供了有效手段。这4种抽象数据类型与集合论和图论的关系如表3-10所示。

表3-10 HDBS与集合论概念和图论概念之间的联系

HDBS理论	集合论	图 论
类别 Class	集 Set	超图的边缘
目标 Object	要素 Element	超图的端点
属性 Attribute	特征 Property	端点赋值或边缘赋值
关联 Relation	相关关系 Relationship	联接

3.5.2 超图数据组织

HDBS根据类别、物体、属性和关系这4种抽象类型，定义了6种表示地理现象的基本单元：类别、类别属性、物体、物体属性、类别联系和物体联系。HDBS根据这6种单元来组织超图数据。

（1）类别：类别是同类目标的集合，它们具有某些相同的性质并可能表示成相同的关系。在超图中用环绕着集合中某些元素的一条边和一个代表类别名字的特殊顶（端）点来表示。

（2）类别属性：类别属性即类别的性质，是由特殊顶点表示的评价值。

(3) 物体：即类别的元素。一个类别中的物体个数不受限制。

(4) 物体属性：物体属性即物体的性质。

(5) 类别联系：类别联系表明同一类别或不同类别物体之间的联系。在两个类别之间可有多种联系，其中每一种联系表示不同性质的联系，这些联系由在超图类别间引绘的弧线来表示。

(6) 物体联系：物体联系表示物体之间实际存在的联系，可以描述特定的现象。

上述两类联系可以是层次性的，也可以是非层次性的。①类别间的层次联系：一个类别可有一个双亲类别，也可有一个或多个子女类别（或空），这是一种树结构，它的结点是处于不同层次的类别。每个类别是把与该类别有关元素圈起来的超图的一条边，类别间的层次联系是一种概念上的联系，是一种联系模式。②物体间的层次联系：它是类别间层次联系的具体表现。例如，各条公路是集合"公路"的元素，它们是"公路"类别中的物体，然而各条公路又由各个路段构成，因而就有一个"路段"类别，它是"公路"类别的一个下属类别。

类别之间或物体之间不仅有竖向联系，而且也有横向联系或非层次联系。例如，类别"公路"与"城镇"之间没有隶属关系，但有对等的"连接"与"被连接"的关系。

HDBS的主要优点在于，它能够区别信息的两种基本结构（层次和非层次），并且任何一种复杂的信息结构都能够根据这两种信息结构来建立。超图数据库的设计是以HDBS为基础的，以HDBS的数据链系统（Data String System，DSS）作为数据库中数据的基本单元，并且在Corbett"单元结构"理论的支持下进行建立，即将超图数据链的头、体分开，分别建立0单元库、1单元库和2单元库。在指针快速搜索系统的支持下，进行数据库的自动维护与更新，以及0单元、1单元和2单元的快速搜索，以实现图形及属性数据的直接联系。

有关超图的理论与技术研究仍在发展过程中，其中具体细节详见有关资料。

参 考 文 献

边馥苓. 1996. 地理信息系统原理和方法. 北京:测绘出版社. 100～112

陈述彭,鲁学军,周成虎. 1999. 地理信息系统导论. 北京:科学出版社

崔伟宏. 1995. 空间数据结构研究. 中国科学技术出版社. 39～44, 76～132

龚健雅. 2001. 地理信息系统基础. 北京:科学出版社. 84～102, 116～121

龚健雅. 1993. 整体SIS数据组织与处理方法. 武汉:武汉测绘科技大学出版社

郭达志等. 2002. 地理信息系统原理与应用. 徐州:中国矿业大学出版社. 128～133, 144～149

黄杏元,汤 勤. 1989. 地理信息系统概论. 北京:高等教育出版社

李德仁,龚健雅,边馥苓. 1993. 地理信息系统导论. 北京:测绘出版社

邬伦等. 2001. 地理信息系统——原理、方法与应用. 北京:科学出版社. 52, 142

吴信才. 2002. 地理信息系统原理与方法. 北京:电子工业出版社. 20～42

张 超,陈丙咸,邬 伦. 1995. 地理信息系统. 北京:高等教育出版社

Atsuyuki O., Barry B., Kokichi S. and Sung N. C., 2000. Spatial tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, INC

Martien M., 1998. An Introduction to The Theory of Spatial Object Modelling for GIS. Talylor & Francis Ltd

Michael N. D., 2000. Fundamentals of Geographical Information Systems. John Wiley & Sons, INC

Paul A. L., Michael F. G., David J. M. and David W. R., 1999. Geographical Information Systems (I): Principles And

Technical. John Willey & Sons, INC

Paul A. L., Michael F. G., David J. M. and David W. R., 1999. Geographical Information Systems (II): Management Issues And Applications. John Willey & Sons, INC

Peter A. B., 1986. Principles of GIS for Land Resources Assessment. Oxford: Clarendon Press

Peter A. B. and Rachael A. M., 1998. Principles of Geographical Information Systems. Oxford University Press

Lee Y. C., Li Z. L. and Li Y. L. 2000. Taxonomy of space tessellation. J. of Photogrammetry & Remote Sensing. 55:139 149

Donna J. P. A conceptual framework and comparison of spatial data models. Cartographica. 21:66－113

第四章 3D GIS空间数据模型

尽管一些GIS软件已经具有2.5D表面建模和3D景观可视化能力，但还不是真3D的。所谓真3D GIS，它是将3D空间坐标（x，y，z）作为独立参数来进行空间实体对象的几何建模，其数学表示为：$F=f(x, y, z)$，因而所建立的模型不仅可以实现3D可视化，还可以进行3D空间分析。目前，真3D GIS的理论研究与产品开发都处于探索与试验阶段，众多学者（包括地理、测绘、计算机科学等领域的学者）对此作了深入的研究，取得了一定的成果，但目前还未开发出一个真正基于3D模型的GIS系统（或称为真3D GIS，Real-3D GIS）。这使得矿山、地质、海洋、环境、城市、气象等众多领域的3D原始数据不能得到充分利用，3D空间信息处理受到严重制约。真3D地质模拟、地面地下工程的空间整合分析以及3D动态模拟等问题，已成为GIS的技术前沿和攻关热点。本章将在对准-3D GIS空间数据模型进行总结分析的基础上，结合3D地学构模对真-3D空间构模、3D拓扑描述与维护等进行讨论，并介绍基于广义三棱柱（GTP）的3D地学构模原理与实例。

§4.1 空间维数与空间构模

近10年来，地学领域（作者将其理解为地理Geography、地质Geology、地球物理Geophysics和大地测量Geodesy等地学相关学科的统称，因其英文名称之前缀均为Geo-）关于3D空间信息的研究与日俱增，形成了两大并行发展的支流：一是3D地理信息系统（3D GIS），二是3D地学模拟系统（3D Geosciences Modeling System，3D GMS）。

4.1.1 3D GIS与3D GMS之比较

3D GIS是随着计算机可视化技术的发展和2D GIS的成熟，而在20世纪90年代初开始为人们所关注。2D GIS的本质是将3D现实世界中的地物与地理现象投影到某一2D平面（通常为XY平面）上进行表达，虽然简化了空间信息理解与表达的过程，却损失了空间信息量（尤其是高程信息Z和3D拓扑空间信息），是以牺牲空间信息的真实性和完整性为代价的。3D GIS正是要针对2D GIS的这一本质缺陷，试图直接从3D空间的角度去理解和表达现实世界中的地物、地理现象及其空间关系。

3D GMS则是随着科学可视化技术和地质信息计算机模拟技术的发展，而几乎在20世纪90年代初的同期开始为人们所重视。传统的地质信息的模拟与表达方式主要有两种，其一是采用平面图和剖面图进行表达（如底板等高线图、采掘工程平面图、地质剖面图、钻孔剖面图等），其实质也是将3D地质环境中地层、矿体与地质现象投影到某一2D平面（XY平面、XZ平面或YZ平面）上进行表达；其二是采用透视和轴侧投

影原理，将3D地质环境中的地层、矿体与地质现象进行透视制图，或投影到两个以上的平面上进行组合表达，以增强3D视觉效果，提高人们的3D理解水平。这两种方式同样存在空间信息的损失与失真问题，而且制图过程繁杂，信息更新困难。3D GMS正是针对传统的地质信息模拟与表达方法的不足和缺陷，借助于计算机和科学可视化技术，直接从3D数字空间的角度去理解和表达地质体与地质环境。

因此，3D GIS与3D GMS在许多方面有相似之处，也有不同之处，如表4-1所示。3D GIS以地球表面及以上为其研究对象，也就是通常所说的地理空间（geographical space）；而3D GMS则以地球表面及以下为其研究对象，也就是通常所说的地质空间（geological space）。两者以地球表面为界，分别处理与地球相关的信息（Geo-information）的不同部分，构成对整个地球系统的统一描述与表达，因而可以统称为地学（Geosciences）中的地球信息学（Geo-informatics）。

表4-1　3D GIS与3D GMS的比较

项目 \ 比较	不同之处		相同之处
	3D GIS	3D GMS	
研究对象	地球表面及以上	地球表面及以下	以地球表面为界
数据来源	大地测量、工程测量、摄影测量与遥感	地质勘探、地球物理、矿山测量、地质解译	规划设计数据、统计分析数据
空间参照	大尺度时球面坐标	大尺度时球体坐标	一般为（X，Y，Z）坐标
空间构模	以面元构模为主	以体元构模为主	面元、体元可混合使用
拓扑描述	不可或缺	开始研究	基于要素或基于对象
空间量算	非以体积计算为主	以体积计算为主	方位、面积计算
空间分析	以可视分析为主	以包含、临近分析为主	其他空间分析
应用领域	城市景观、立体交通、管网规划、军事、环境	地下空间、矿山地质、土木工程、海洋、水文	地球整体、地上下整合

4.1.2　空间维数讨论

在GIS领域，人们通常将以平面制图和平面分析为主要功能的GIS，称为2D GIS；将增加了高程信息，可以构建数字高程模型（DEM）或数字地形模型（DTM）的GIS，称为二维半或2.5D GIS。近年，随着GIS和GMS向3D方向发展，空间维数一直是一个存在分歧和争议的问题（Turner A. K.，1991，1998），出现了一些新的名词，如2.75D GIS、假-3D GIS或伪-3DGIS（pseudo-3D GIS）（Mei-Po K.，2000）、3D城市模型（3D CM）（Monika R. et al.，1997；孙敏等，2000）、假3D-GMS、真-3D GIS和真-3D GMS（Wu L.X.，2002；吴立新等，2002，2003a，2003b）。实际上，无论是2.75D GIS还是假-3D GIS、假-3D GMS，或是3D CM，它们与真-3D GIS或真-3D GMS的本质区别在于：前者只描述3D空间实体的外部轮廓，不表达其内部属性；而后者则是以体元的方式在描述3D空间实体的外部轮廓的同时，还表达其内部属性。为统一认识，作者建议采用一个较中性的词——"准"（quasi），将2.75D GIS、假-3D

GIS 和伪-3D GIS 统称为准-3D GIS，将假-3D GMS 称为准-3D GMS（吴立新等，2003a）。

可以从数学模型和空间构模技术来比较不同空间维数的 GIS 及 GMS 的差异，其结果如表 4-2 所示。

表 4-2　空间维数分析

类别 / 比较	2D GIS	2.5 D GIS	准 3D GIS 与准 3D GMS	真 3D GIS 与真 3D GMS
数学模型	$F=f(x,y)$	$F=f(x,y)$ 和 $Z=f(x,y)$	$F=f(x,y,z_i)$	$F=f(x,y,z_i)$
高程特征	无高程信息	高程作为点的属性	一对(x,y)的z有多值	一对(x,y)的z有多值
属性特征	平面抽象	表面抽象	无体内属性	有体内属性
构模方式	2D 矢量或栅格	2D 矢量或栅格	面元构模	3D 矢量、体元构模
典型实例	电子地图	DEM、DTM	3D 城市模型	地下 TEN、GTP 模型

（1）2D GIS 仅是将平面坐标(x,y)作为独立参数来进行平面对象的几何建模，其数学表示为：$F=f(x,y)$；其空间构模方式为矢量或栅格方式。

（2）2.5D GIS 中高程仅作为一个属性值（Ferenc S.，1994），数学表示为：$F=f(x,y)$与 $Z=f(x,y)$，或表示为 $F=f[x,y,z(x,y)]$；其空间构模方式为矢量或栅格方式，典型的 DTM 一般采用 TIN 或 Grid 方式。

（3）准-3D GIS 是在 DTM 之上叠加地面建筑设施的 3D 造型，如 CSG 模型（李清泉，1998；孙敏，陈军，2000）、TEN 模型（Joe B.，1991；Victor J. D. et al.，1993；Pilout M. et al.，1994），通常两者是分离的；而准-3D GMS，则是基于面元构模，因而无法处理垂向上两个以上的采样值或界面，无法进行任意剖面切割。两者均无法进行 3D 空间查询和 3D 空间分析等真-3D 操作，更无法进行一体化空间操作和空间分析。

（4）真-3D GIS，则是将 3D 空间坐标(x,y,z)作为独立参数来进行空间实体对象的几何建模，其数学表示为：$F=f(x,y,z)$；且必须采用 3D 矢量构模或 3D 体元构模方式。因而所建立的模型不仅可以实现真 3D 可视化，还可以进行 3D 空间分析。

4.1.3　空间构模技术分类

3D 空间构模方法研究是目前 3D GIS 领域以及 3D GMS 领域研究的热点问题。许多专家学者在此领域做了有益的探索。地质、矿山领域的一些专家学者，围绕矿床地质、工程地质和矿山工程问题，对 3D GMS 的空间构模问题进行了卓有成效的理论与技术研究，加拿大、澳大利亚、英国、南非等国还相继推出了一批在矿山和工程地质领域得到推广应用的 3D GMS 软件。

过去十来年中，研究提出了 20 余种空间构模方法。若不区分准-3D 和真-3D，则可以将现有空间构模方法归纳为基于面模型（facial model）、基于体模型（volumetric

model）和基于混合模型（mixed model）的3大类构模体系，如表4-3所示。

表4-3 3D空间构模法分类

面模型(facial model)	体模型(volumetric model)		混合模型(mixed model)
	规则体元	非规则体元	
不规则三角网(TIN)	结构实体几何(CSG)	四面体格网(TEN)	TIN-CSG混合
格网(Grid)	体素(Voxel)	金字塔(Pyramid)	TIN-Octree混合或Hybrid模型
边界表示模型(B-Rep)	八叉树(Octree)	三棱柱(TP)	Wire Frame-Block混合
线框(Wire Frame)或相连切片(Linked Slices)	针体(Needle)	地质细胞(Geocellular)	Octree-TEN混合
断面序列(Series Sections)	规则块体(Regular Block)	非规则块体(Irregular Block)	
断面-三角网混合(Section-TIN mixed)		实体(Solid)	
多层DEMs		3D Voronoi图	
		广义三棱柱(GTP)	

§4.2 3D空间构模原理

4.2.1 基于面模型的准3D空间构模

基于面模型的构模方法侧重于3D空间实体的表面表示，如地形表面、地质层面、构筑物（建筑物）及地下工程的轮廓与空间框架。所模拟的表面可能是封闭的，也可能是非封闭的。基于采样点的TIN模型和基于数据内插的Grid模型，通常用于非封闭表面模拟；而B-Rep模型和Wire Frame模型通常用于封闭表面或外部轮廓模拟。Section模型、Section-TIN混合模型及多层DEM模型通常用于地质构模。通过表面表示形成3D空间目标轮廓，其优点是便于显示和数据更新，不足之处由于缺少3D几何描述和内部属性记录而难以进行3D空间查询与分析。

（1）TIN与Grid模型（Christopher C. M. et al., 1978；Nicolas S. et al., 1991；Victor J. D., 1993）：有很多方法可以用来表达表面，如等高线（Contour）模型、Grid模型（如图4-1所示）、TIN模型等。最常用的表面构模技术是基于实际采样点构造TIN。TIN方法将无重复点的散乱数据点集按某种规则（如Delaunay规则）进行三角剖分，使这些散乱点形成连续但不重叠的不规则三角面片网，并以此来描述3D物体的表面；而Grid模型则是考虑到采样密度和分布的非均匀性，经内插处理后形成规则的平面分割网格。这两种表面模型一般用于地形表面构模，也可用于层状矿床构模。对于层状矿床，一般先生成各岩层接触界面或厚度在模型域上的表面模型，然后根据岩层间的截割和切错关系通过“修剪”、“优先级次序覆盖”、算术和逻辑运算方法对各岩层接触面或厚度进行精确修饰（如EAGLES系统）。

（2）边界表示（B-rep）模型：通过面、环、边、点来定义形体的位置和形状，边界

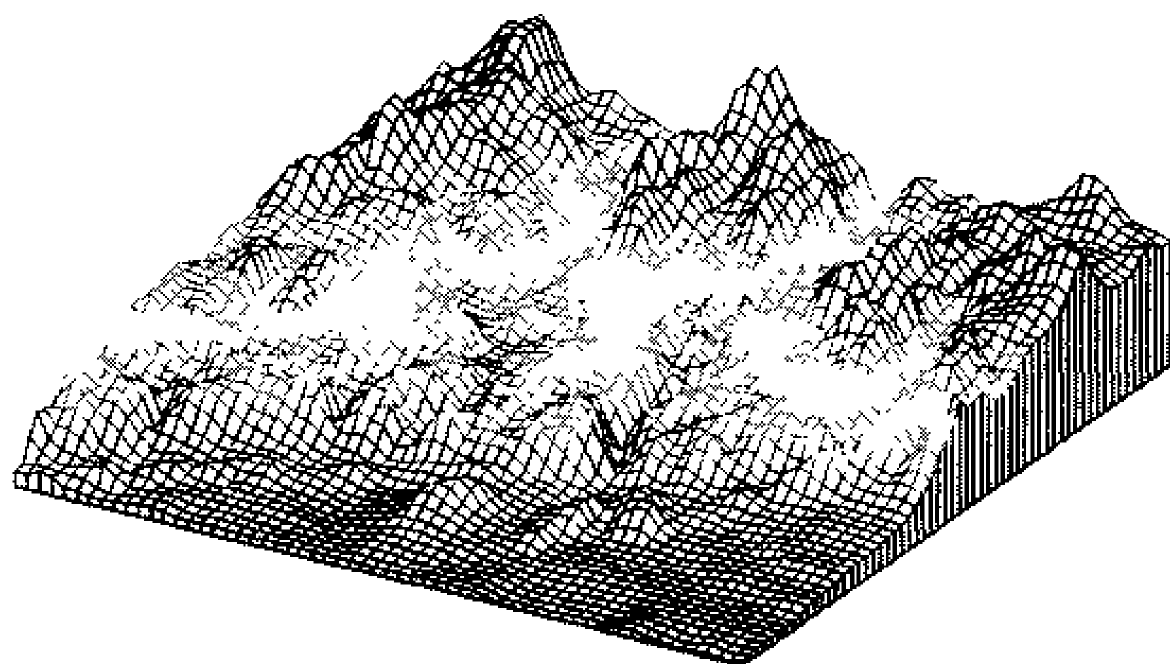


图 4-1 基于 Grid 的地形模拟

线可以是平面曲线，也可以是空间曲线。例如一个长方体由 6 个面围成，对应有 6 个环，每个环由 4 条边界定，每条边又由两个端点定义，如图 4-2 所示。其特点是：详细记录了构成物体形体的所有几何元素的几何信息及其相互连接关系，以便直接存取构成形体的各个面、面的边界以及各个顶点的定义参数，有利于以面、边、点为基础的各种几何运算和操作。边界表示构模在描述结构简单的 3D 物体时十分有效，但对于不规则 3D 地物则很不方便，且效率低下。

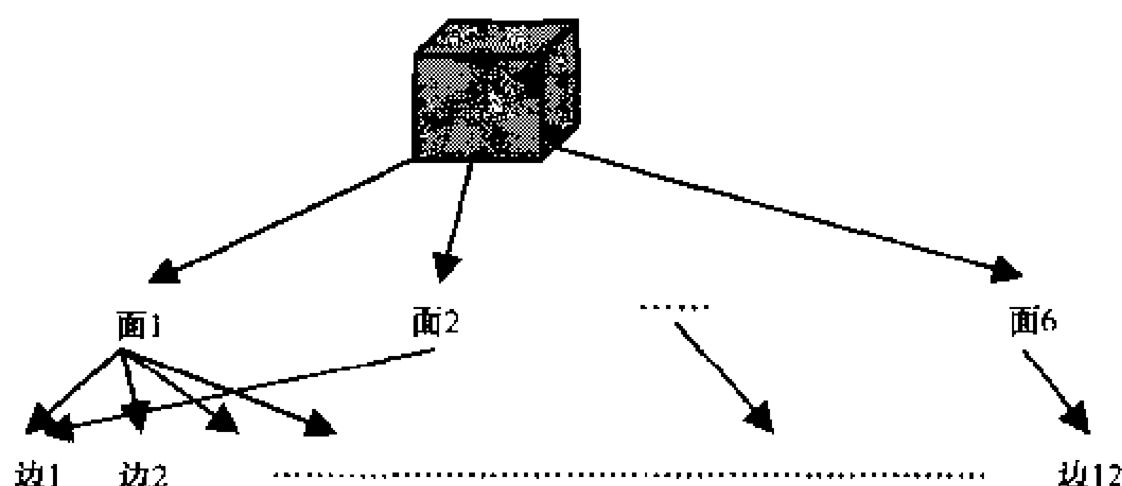
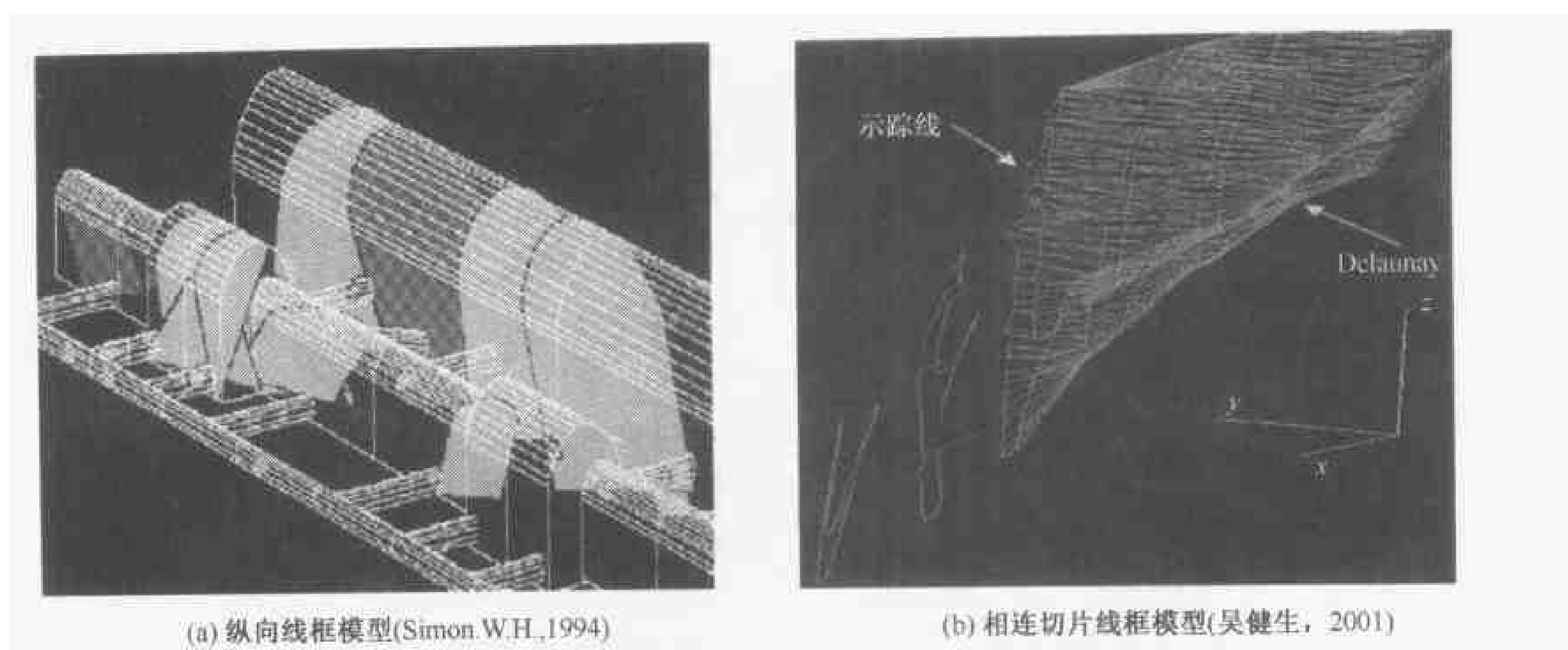


图 4-2 B-rep 构模示例

(3) 线框（Wire Frame）模型（赫尔丁 S. W.，1989；Simon W. H.，1994；许斌等，1994）：线框构模技术实质是把目标空间轮廓上两两相邻的采样点或特征点用直线连接起来，形成一系列多边形；然后把这些多边形面拼接起来形成一个多边形网格来模拟地质边界或开挖边界（如图 4-3 所示）。某些系统则以 TIN 来填充线框表面，如 DataMine。当采样点或特征点沿环线分布时，所连成的线框模型也称为相连切片（Linked Slices）模型，或连续切片模型（如图 4-3b 所示）。

(4) 序列断面（Series Sections）模型（阿列尼切夫 B. M. 等，1995；朱小弟，2001a）：断面构模技术实质是传统地质制图方法的计算机实现，即通过平面图或剖面图来描述矿床，记录地质信息，如图 4-4 所示。其特点是将 3D 问题 2D 化，简化了程序设计；同时在地质描述上它也是最方便、使用性最强的一种构模技术；但它在矿床的表



(a) 纵向线框模型(Simon.W.H.,1994)　　(b) 相连切片线框模型(吴健生，2001)

图 4-3　基于线框的地下构模

达上是不完整的，断面构模难以完整表达 3D 矿床及其内部结构，往往需要通过其他构模方法配合使用。此外，由于采用的是非原始数据而存在误差，其构模精度一般难以满足工程要求。

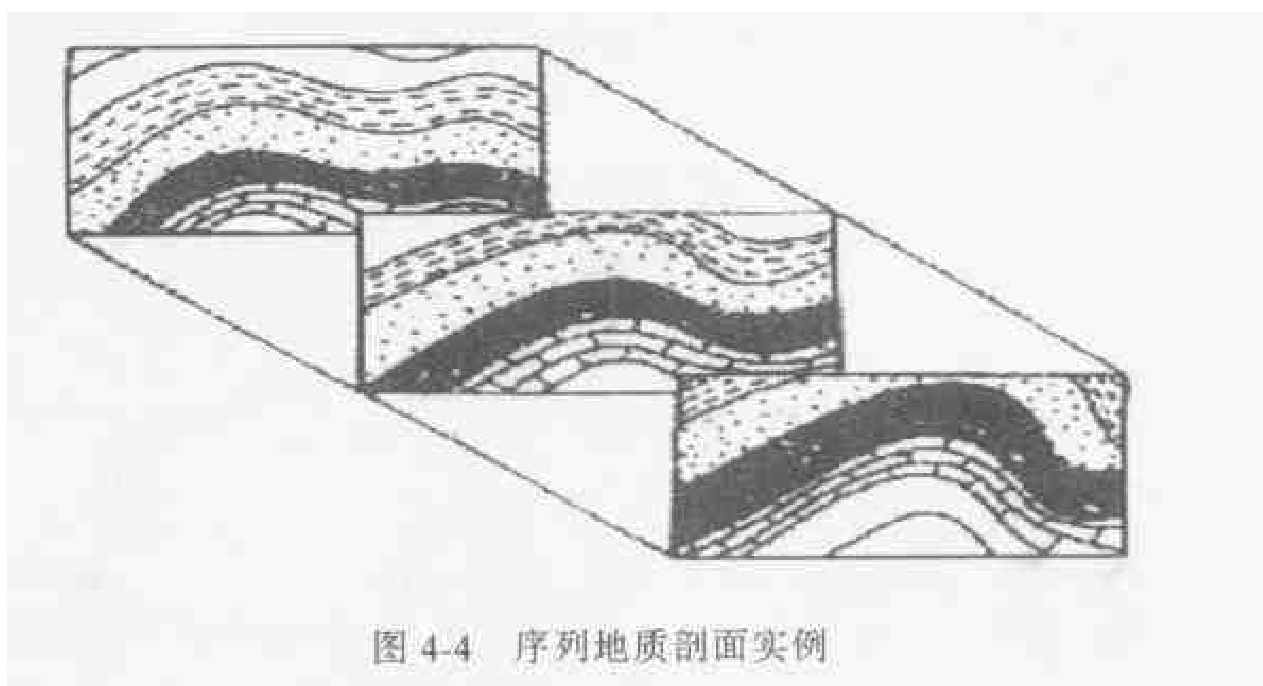


图 4-4　序列地质剖面实例

(5) 断面-三角网混合模型（朱小弟，2001b）：在二维的地质剖面上，主要信息是一系列表示不同地层界线的或有特殊意义的地质界线（如断层、矿体或侵入体的边界），每条界线赋予属性值，然后将相邻剖面上属性相同的界线用三角面片（TIN）连接，这样就构成了具有特定属性含义的 3D 曲面，如图 4-5 所示。其构模步骤为：①剖面界线赋值；②二维剖面编辑；③相邻剖面连接和④3D 场景的重建。

(6) 多层 DEM 构模（赵树贤，1999）：首先基于各地层（尤其是控制性地层或关键地层）的界面点按 DEM 的方法对各个地层进行插值或拟合，然后根据各地层的属性对多层 DEM 进行交叉划分处理，形成空间中严格按照岩性（或土壤性质）为要素进行划分的 3D 地层模型的骨架结构。在此基础上，引入地下空间中的特殊地质现象、人工构筑物等点、线、面、体对象，完成对 3D 地下空间的完整剖分，如图 4-6 所示。

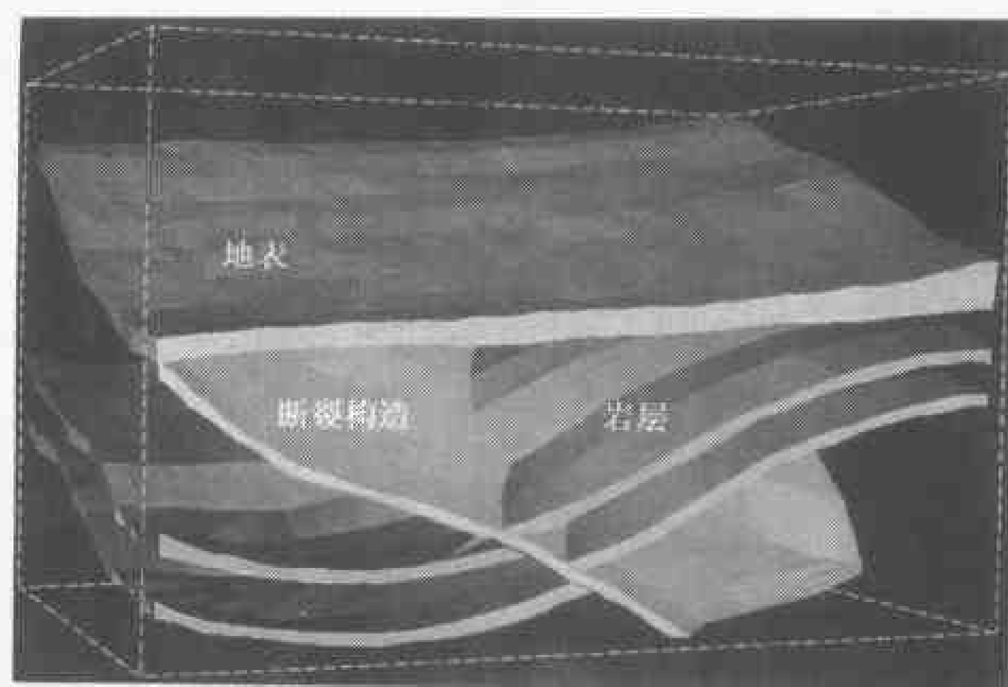
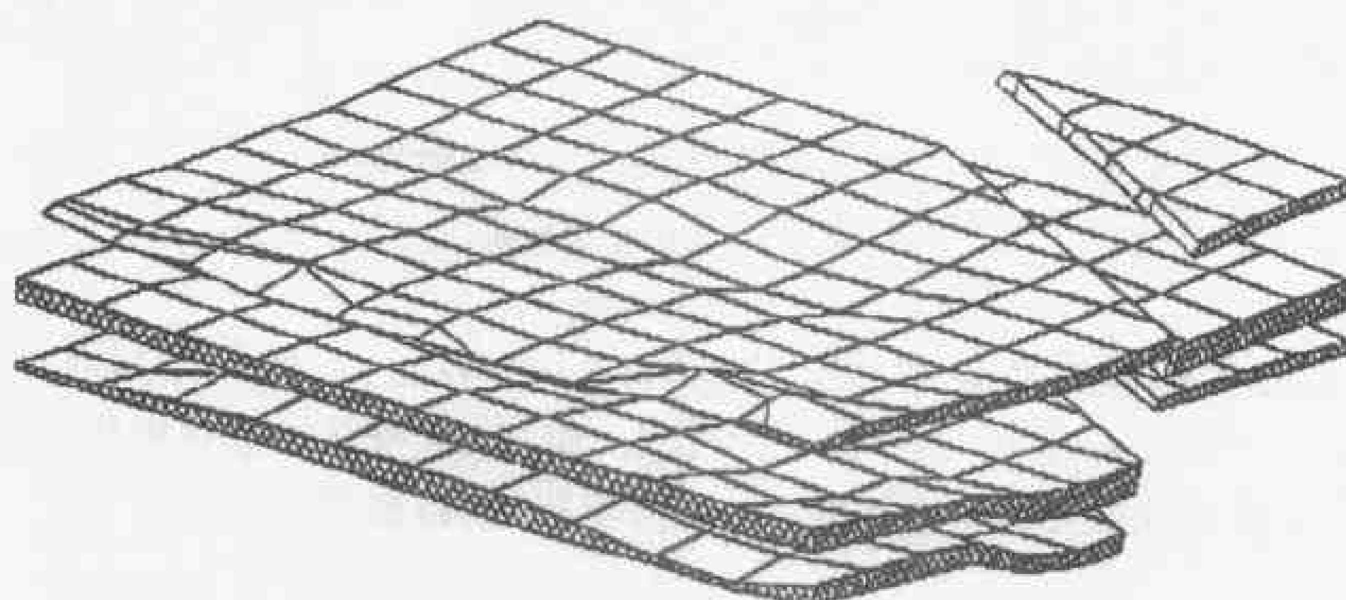


图 4-5 基于断面-三角网的混合构模实例
（朱小弟，2001b）


图 4-6 多层 DEM 构模实例
（赵树贤，1999）

4.2.2 基于体模型的真 3D 空间构模

真 3D 地学模拟、地面与地下空间的统一表达、陆地海洋的统一建模、3D 拓扑描述、3D 空间分析、3D 动态地学过程模拟等问题，已成为地学与信息科学的交叉技术前沿和攻关热点。

体模型基于 3D 空间的体元分割和真 3D 实体表达，体元的属性可以独立描述和存储，因而可以进行 3D 空间操作和分析。体元模型可以按体元的面数分为四面体（Tetrahedral）、六面体（Hexahedral）、棱柱体（Prismatic）和多面体（Polyhedral）共 4 种类型，也可以根据体元的规整性分为规则体元和非规则体元两个大类。规则体元包括 CSG-tree、Voxel、Octree、Needle 和 Regular Block 共 5 种模型，如图 4-7 所示。规则体元通常用于水体、污染和环境问题构模，其中 Voxel、Octree 模型是一种无采样约束的面向场物质（如重力场、磁场）的连续空间的标准分割方法，Needle 和 Regular Block

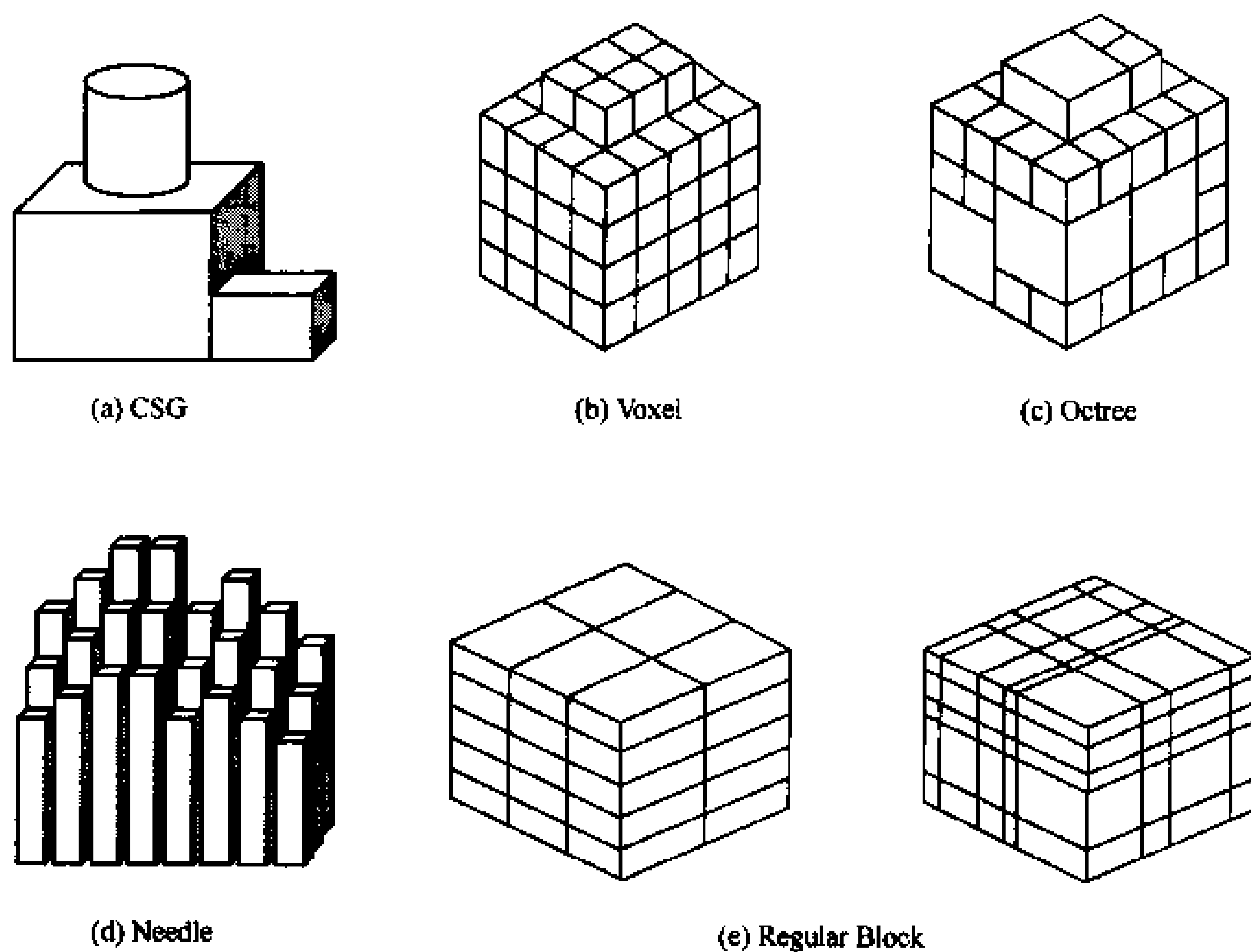


图 4-7　规则体元模型

可用于简单地质构模。非规则体元包括 TEN、Pyramid、TP、Geocelluar、Irregular Block、Solid、3D-Voronoi 和 GTP 共 8 种模型，如图 4-8 所示。非规则体元均是有采样约束的、基于地质地层界面和地质构造的面向实体的 3D 模型。

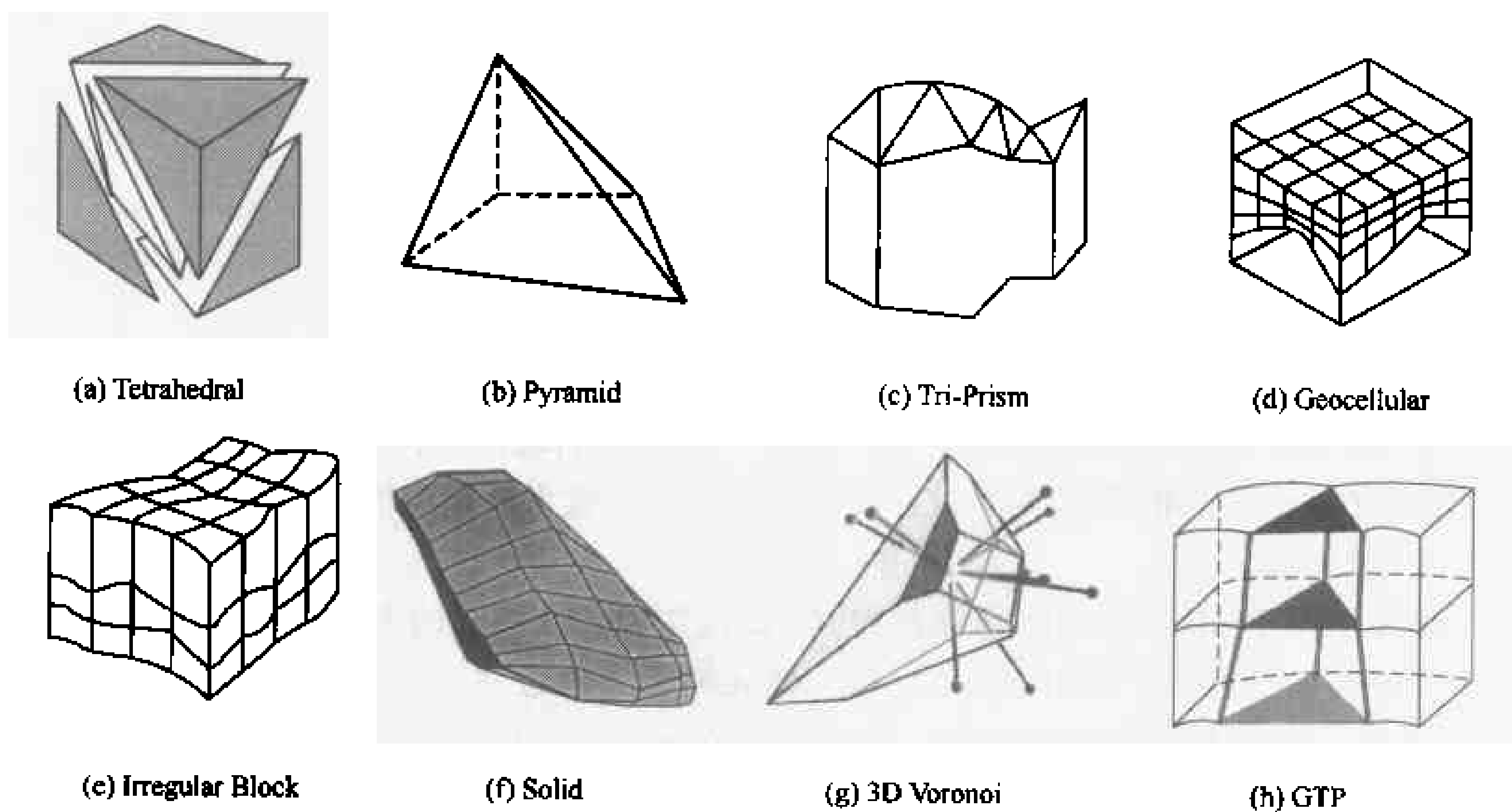


图 4-8　非规则体元模型

1. 基于规则体元的构模

(1) 结构实体几何（CSG）构模（李清泉，1998；孙敏等，2000）：首先预定义好一些形状规则的基本体元，如立方体、圆柱体、球体、圆锥及封闭样条曲面等，这些体元之间可以进行几何变换和正则布尔操作（并、交、差），由这些规则的基本体元通过正则操作来组合成一个物体。生成的 3D 物体可以用 CSG 树来表示。CSG 构模在描述结构简单的 3D 物体时十分有效，但对于复杂不规则 3D 地物尤其是地质体则很不方便，且效率也大大降低。CSG 的构模原理如图 4-9 所示。

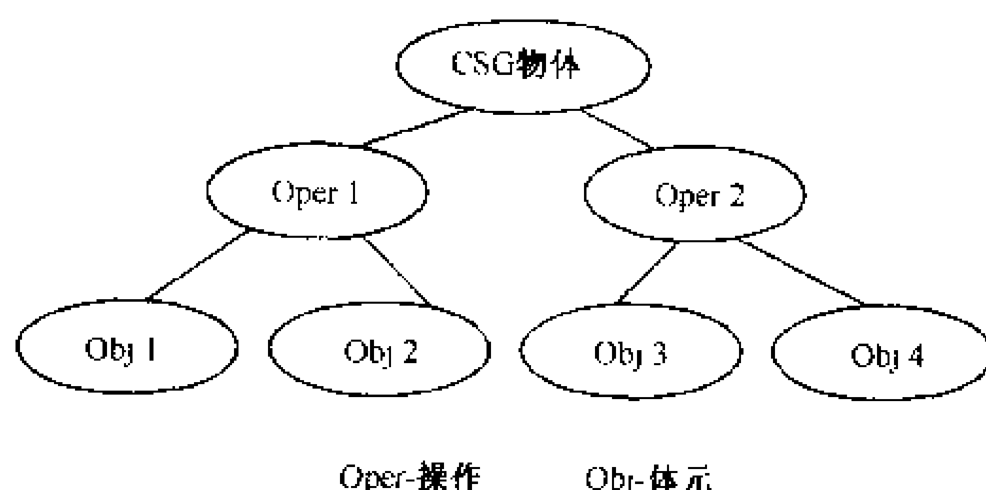


图 4-9 CSG 的构模原理

(2) 3D 体素（Voxel）构模（Simon W. H.，1994）：该模型的实质是 2D Grid 模型的 3D 扩展，即以一组规则尺寸的 3D 体素（$a=b=c$）来剖分所要模拟的空间。基于 Voxel 的构模法有一个显著优点，就是在编制程序时可以采用隐含的定位技术，以节省存储空间和运算时间。该模型虽然结构简单，操作方便，但表达空间位置的几何精度低，且不适合于表达和分析实体之间的空间关系。当然，通过缩小 Voxel 的尺寸，可以提高构模精度，但空间单元数目及储量将成三次方增长。

(3) 八叉树（Octree）构模（韩国建等，1992；Homer H. et al.，1998；肖乐斌等，1998；边馥苓等，2000）：类似于 2D GIS 中的四叉树之于 Grid，Octree 模型实质上是对 Voxel 模型的压缩改进。该方法将 3D 空间区域分成 8 个象限，且在树上的每个节点处存储 8 个数据元素。当象限中所有体元的类型相同时（即为均质体），该类型值存入相应的节点数据元素中。非均质象限再进行象限细分，并由该节点中的相应数据元素指向树中的下一个节点，如此细分直到每个节点所代表的区域都是均质体为止，如图 4-10 所示。Octree 模型在医学、生物学、机械学等领域已得到成功应用，但在矿床地质构模中有较大的局限性。基于八叉树，肖乐斌等（1998）提出了四层矢量化八叉树层次结构，边馥苓等（2000）提出了面向目标的数据结构。

(4) 针体（Needle）构模：Needle 模型的原理类似于结晶生长过程，用一组具有相同截面尺寸的不同长度或高度的针状柱体对某一非规则 3D 空间、3D 地物或地质体进行空间分割，用其集合来表达该目标空间、3D 地物或地质体。

(5) 规则块体（Regular Block）构模（Simon W. H.，1994）：块体构模技术的研究和应用始于 60 年代初，是一种传统的地质构模方法。60 年代和 70 年代开发的一些计算机系统采用这种构模技术，比较典型的有 RTZ 公司开发的 OBMS 和 OPDP 系统，

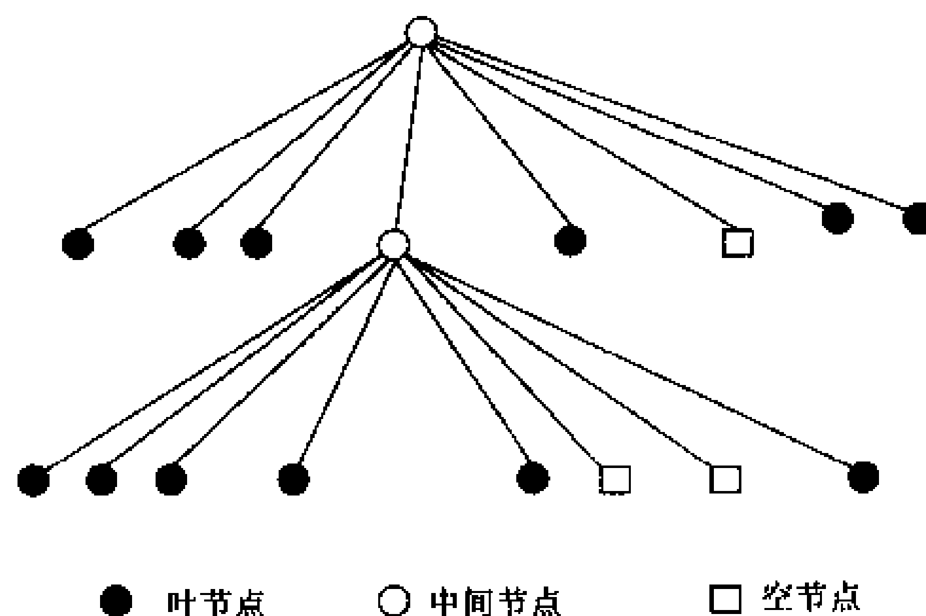


图 4-10 八叉树构模原理
（韩国建等，1992）

Control Data 公司的 MINEVAL 系统和 Minetec 公司的 MEDS 系统。这类构模技术是把要建模的空间分割成规则的 3D 立方网格，称为 Block，每个块体在计算机中的存储地址与其在自然矿床中的位置相对应，每个块体被视为均质同性体，由克立格法、距离加权平均法或其他方法确定其品位或岩性参数值。该模型用于属性渐变的 3D 空间（如侵染状金属矿体）构模很有效，对于有边界约束的沉积地层、地质构造和开挖空间的构模则必须不断降低单元尺寸，从而引起数据急速膨胀。解决方法是在边界区域进行局部的单元细化（如 DataMine 系统）。

2. 基于非规则体元的构模

（1）四面体格网（TEN）构模（Joe B.，1991；Pilout M. et al.，1994；Victor J.D. et al.，1993）：TEN 模型是在 3D Delaunay 三角化研究的基础上提出的，是一个基于点的 TEN（Tetrahedral Network）的 3D 矢量数据模型。其基本思路是对 3D 空间中无重复的散乱点集用互不相交的直线将空间散乱点两两连接形成三角面片，再由互不穿越的三角面片构成 TEN。其中四面体（Tetrahedral）都是以空间散乱点为其顶点，且每个四面体内不含有点集中的任一点。TEN 构模时，四面体内点的属性可由插值函数得到，其中插值函数的参数由四个顶点的属性决定，因此，经过四面体剖分插值后，可以得到空间的 3D 数据信息。TEN 虽然可以描述实体内部，但不能表示 3D 连续曲面，而且用 TEN 来生成 3D 空间曲面也较为困难，算法设计较复杂。Pilout M. et al.（1994）提出的 TEN 数据模型如图 4-11 所示。

（2）金字塔（Pyramid）模型：类似于 TEN 模型，只不过是用 4 个三角面片和 1 个四边形封闭形成的金字塔状模型来实现对空间数据场的剖分。由于其数据维护和模型更新困难，一般很少采用。

（3）三棱柱（Tri-Prism，TP）构模（Simon W. H.，1994；戴吾蛟等，2001；张煜等，2001）：TP 模型是一种较常采用的简单的 3D 地学空间构模技术。张煜等（2001）给出了 TP 体元的定义，同时给出了相关切割和剖分算法，还列举了基于该模型的数字地层模型的相关应用。由于 TP 模型的前提是三条棱边相互平行，因而不能基于实际的

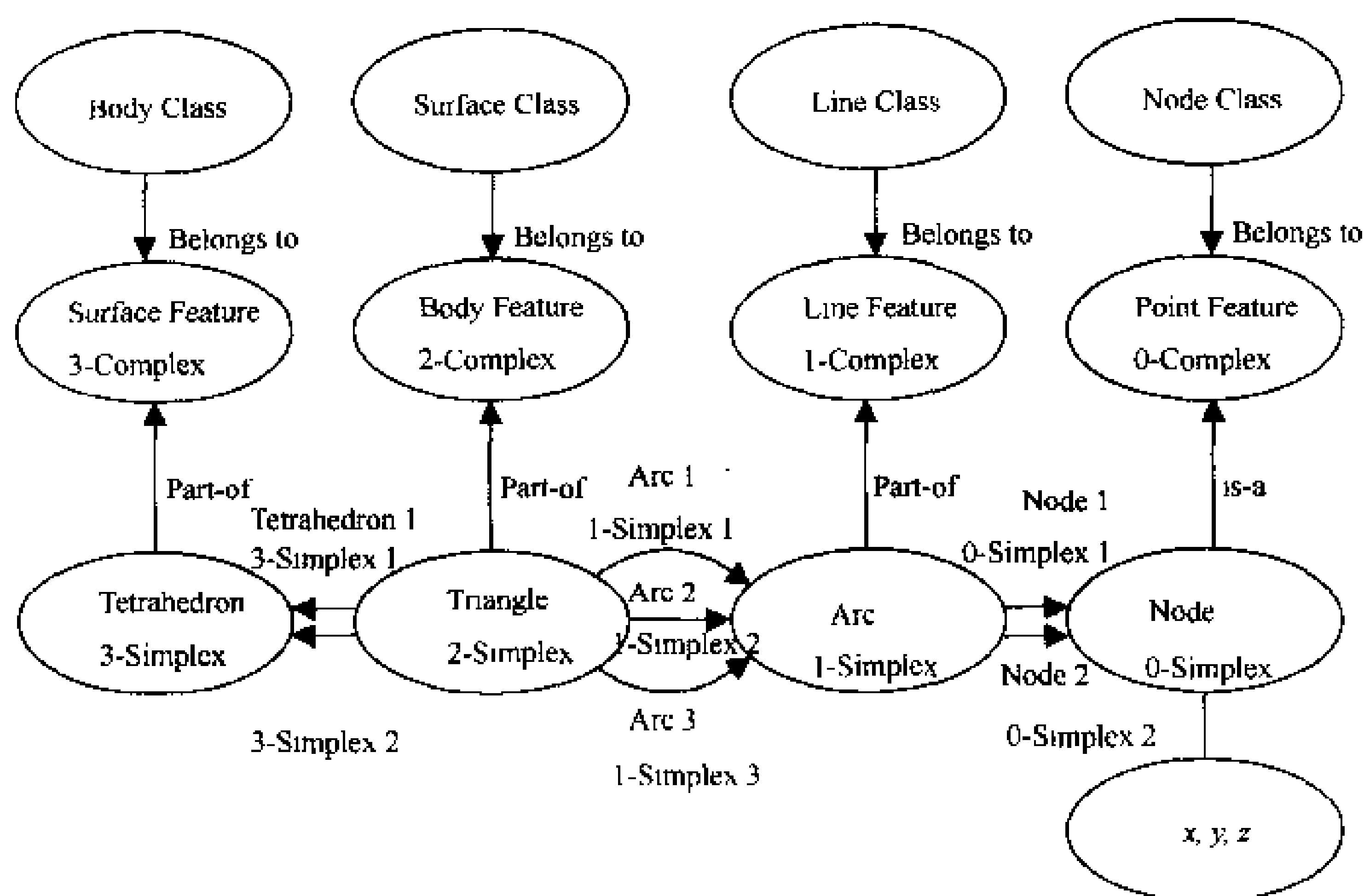


图 4-11 TEN 数据模型

(Pilout M. et al., 1994)

偏斜钻孔来构建真 3D 地质，也难以处理复杂地质构造。戴吾蛟等（2001）则以不规则 TP 为基本单元，讨论了不规则 TP 网络模型（TPN）的数据结构、拓扑建立、拓扑检查和空间插值问题，但对其地学应用方面缺乏深入讨论。

（4）地质细胞（Geocellular）模型：其实质是 Voxel 模型的变种，即在 XY 平面上仍然是标准的 Grid 剖分，而在 Z 方向则依据数据场类型或地层界面变化进行实际划分，从而形成逼近实际界面的 3D 体元空间剖分。

（5）非规则块体（Irregular Block）构模（Simon W. H.，1994）：非规则块体与规则块体的区别在于：规则块体 3 个方向上的尺度（a、b、c）互不相等，但保持常数（如 OBMS 系统）；而非规则块体 3 个方向上的尺度（a、b、c）不仅互不相等，且不为常数。非规则块体构模法的优势是可以根据地层空间界面的实际变化进行模拟，因而可以提高空间构模的精度。

（6）实体（Solid）构模（Simon W. H.，1994）：该法采用多边形网格来精确描述地质和开挖边界，同时采用传统的块体模型来独立地描述形体内部的品位或质量的分布，既可以保证边界构模的精度，又可以简化体内属性表达和体积计算。以加拿大 Lynx 系统中提供的 3D 元件构模（3D Component Modeling）技术为代表，该技术以用户熟悉的和真实的地质或开挖形态为基础，以交互方式模拟生成由地质分表面（Sub-Surface）或开挖边界面构成的 3D 形体，称作元件（Component）。元件不仅表示一个形体，也表示封闭的体积以及形体中的地质特征（品位或质量等）分布。相邻元件相连成组即为一个地质单元或一个开挖单元。实体构模适合具有复杂内部结构（如复杂断层、褶皱和节理等精细地质结构）的构模，如图 4-12 所示。该法的缺点是人工交互工作量巨大，需要耐心细致的工作。

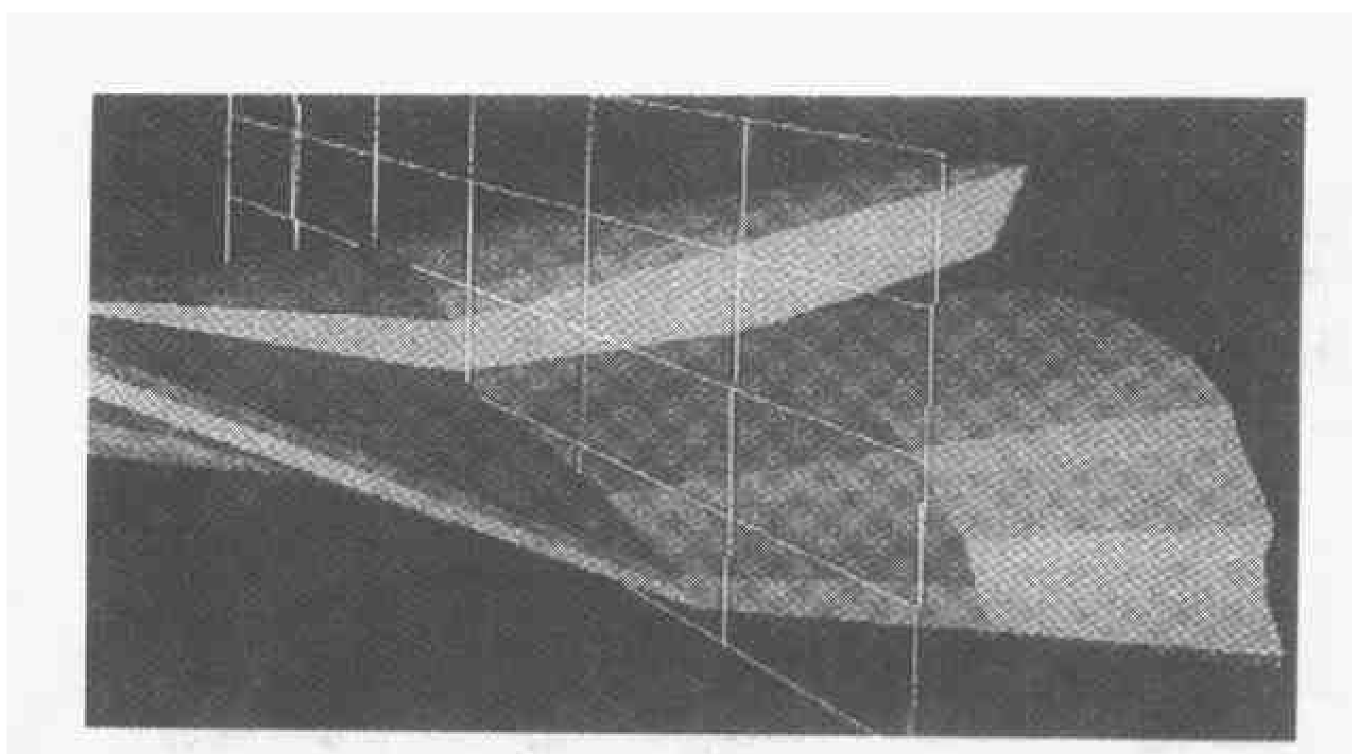


图 4-12 实体构模元件
(Simon W. H., 1994)

(7) 3D Voronoi 图模型：3D Voronoi 图是 2D Voronoi 图的 3D 扩展。其实质是基于一组离散采样点，在约束空间内形成一组面－面相邻而互不交叉（重叠）的多面体，用该组多面体完成对目标空间的无缝分割。3D Voronoi 图模型最早起源于计算机图形学领域；近年，人们开始研究其在地学领域中的可行性，试图在海洋、污染、水体及金属矿体构模方面得到应用。

4.2.3 基于面-体混合的 3D 空间构模

基于面模型的构模方法侧重于 3D 空间实体的表面表示，如地形表面、地质层面等，通过表面表示形成 3D 目标的空间轮廓，其优点是便于显示和数据更新，不足之处是难以进行空间分析。基于体模型的构模方法侧重于 3D 空间实体的边界与内部的整体表示，如地层、矿体、水体、建筑物等，通过对体的描述实现 3D 目标的空间表示，优点是易于进行空间操作和分析，但存储空间大，计算速度慢。混合模型的目的则是综合面模型和体模型的优点，以及综合规则体元与非规则体元的优点，取长补短。

(1) TIN-CSG 混合构模（李清泉，1998；孙敏等，2000）：这是当前城市 3D GIS 和 3D CM 构模的主要方式，即以 TIN 模型表示地形表面，以 CSG 模型表示城市建筑物，两种模型的数据是分开存储的。为了实现 TIN 与 CSG 的集成，在 TIN 模型的形成过程中将建筑物的地面轮廓作为内部约束，同时把 CSG 模型中建筑物的编号作为 TIN 模型中建筑物的地面轮廓多边形的属性，并且将两种模型集成在一个用户界面。这种集成是一种表面上的集成方式，一个目标只由一种模型来表示，然后通过公共边界来连接，因此其操作与显示都是分开进行。

(2) TIN-Octree 混合构模（Hybrid 构模）（Shi，1996，2000）：即以 TIN 表达 3D 空间物体的表面，以 Octree 表达内部结构。用指针建立 TIN 和 Octree 之间的联系。其中 TIN 主要用于可视化与拓扑关系表达。这种模型集中了 TIN 和 Octree 的优点，使拓扑关系搜索很有效，而且可以充分利用映射和光线跟踪等可视化技术。缺点是 Octree

模型数据必须随 TIN 数据的改变而改变，否则会引起指针混乱，导致数据维护困难。

(3) Wire Frame-Block 混合构模（惠勒 A.J. 等，1989）：即以 Wire Frame 模型来表达目标轮廓、地质或开挖边界，以 Block 模型来填充其内部。为提高边界区域的模拟精度，可按某种规则对 Block 进行细分，如以 Wire Frame 的三角面与 Block 体的截割角度为准则来确定 Block 的细分次数（每次可沿一个方向或多个方向将尺寸减半）。该模型实用效率不高，即每一次开挖或地质边界的变化都需进一步分割块体，即修改一次模型。

(4) Octree-TEN 混合构模（李德仁等，1997；李清泉，1998；李清泉等，1998）：随着空间分辨率的提高，Octree 模型的数据量将呈几何级数增加，且八叉树模型始终只是一个近似表示，原始采样数据一般也不保留。而 TEN 模型则可以保存原始观测数据，具有精确表示目标和表示较为复杂的空间拓扑关系的能力。对于一些特殊领域，如地质、海洋、石油、大气等，单一的 Octree 或 TEN 模型很难满足需要，例如在描述具有断层的地质构造时，断层两边的地质属性往往是不同的，需要精确描述。因此，可以将两者结合起来，建立综合两者优点的 Octree-TEN 混合模型，如图 4-13 所示。该模型以 Octree 作整体描述，以 TEN 作局部描述。该混合模型虽然可以解决地质体中断层或结构面等复杂情况的建模问题，但空间实体间的拓扑关系不易建立。

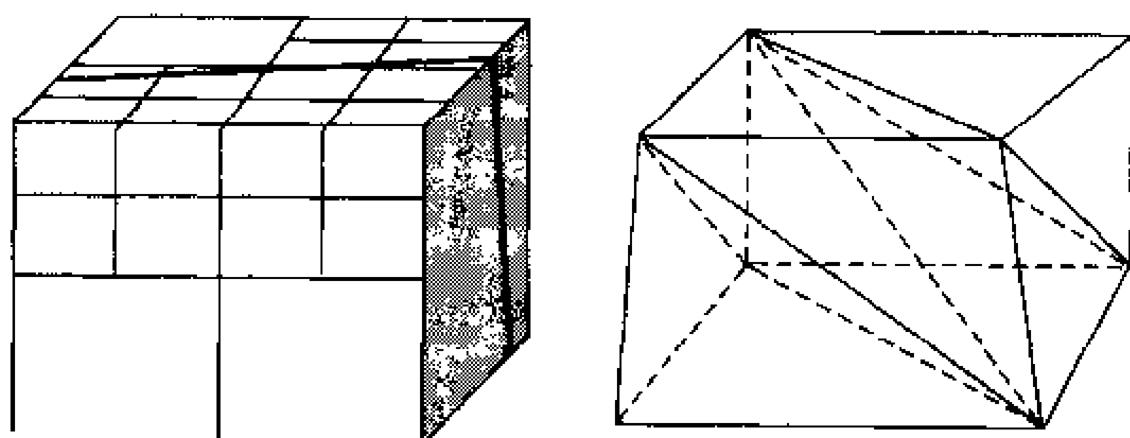


图 4-13 Octree-TEN 混合构模

此外，Shi W. Z. 等（2003）将 3D 空间对象抽象为点、线、面、体（简单体和复杂体）4 个层次，研究了面向对象 3D 模型（OO3D）的概念模型、逻辑模型和形式化描述，并以香港 3DCM 为例进行了实验分析。

§4.3 基于 GTP 的 3D 地学构模

由于 TP 模型的三条棱边相互平行，而且缺乏拓扑描述，只适用于钻孔垂直无偏斜或简单的浅层地基模拟与可视化。TP 模型的应用领域主要是工程地质建模，数据来源是较浅的垂直钻孔数据，基本不存在钻孔偏斜问题，因而可以不考虑钻孔偏斜。而实际上，地质与采矿建模的数据来源一般均来自 100m 以上（甚至 1000m 以上）的深钻，而这种深钻通常存在偏斜，即钻孔轨迹既不垂直也不是直线，而是一条空间曲线，所以不能用 TP 模型来建模。作者针对地质钻孔尤其是深钻的偏斜特点，提出了一种不受三棱柱棱边平行即（钻孔垂直）限制的新的 3D 构模方法，称为类三棱柱（Analogical Tri-Prism, ATP）构模技术（吴立新等，2002；齐安文等，2002），后发展为广义三棱柱

(Generalized Tri-Prism, GTP) 构模技术 (Wu L.X., 2002, 2003), TP 模型仅仅是它的一个特例。

4.3.1 GTP 构模原理

GTP 构模原理是：用 GTP 的上下底面的三角形集合所组成的 TIN 面来表达不同的地层面，然后利用 GTP 侧面的空间四边形面来描述层面间的空间邻接关系，用 GTP 柱体来表达层与层之间的内部实体。

图 4-14 所示为 GTP 构模的基本单元。其中黑粗的柱体代表钻孔，三角形代表 TIN 面。地层被分为两层。由于由上、下不平行的两个 TIN 三角形面和三个侧面空间四边形面所组成的空间单元，与 TP 近似，但又不是 TP（TP 的侧面是平面），故称为 GTP。

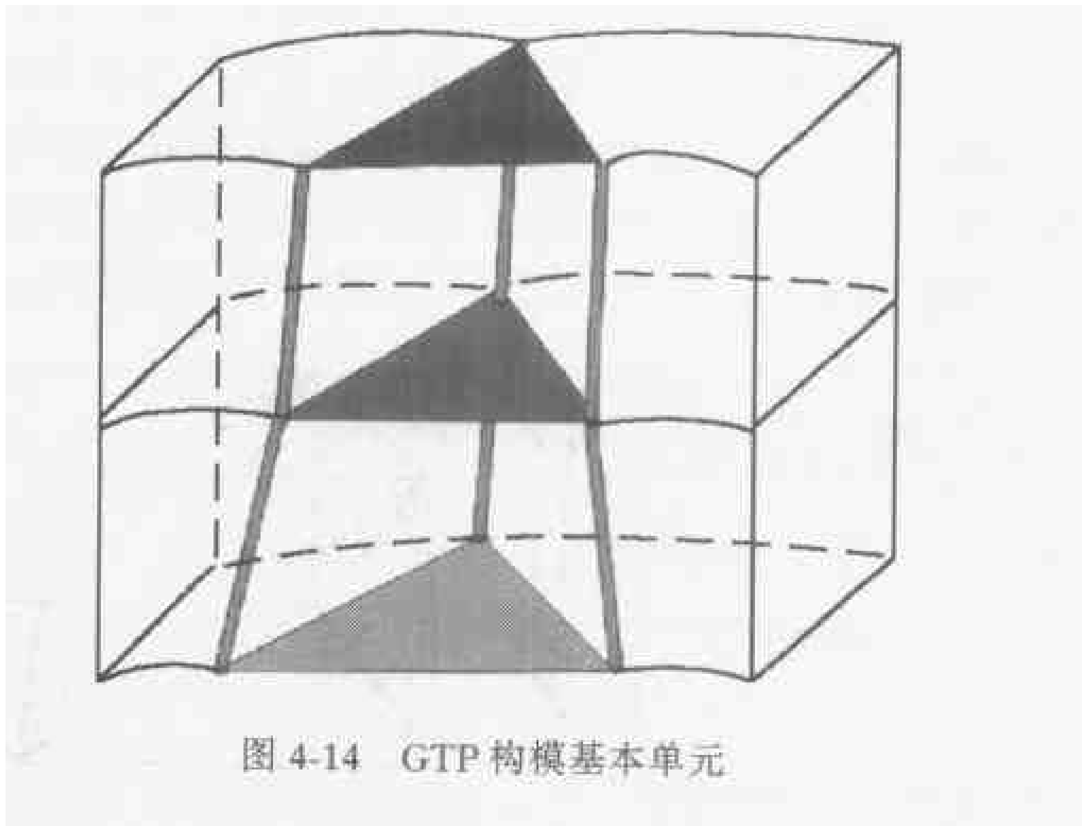


图 4-14 GTP 构模基本单元

如图 4-15 所示，GTP 构模单元有 6 类基本元素组成：结点（Node, P_1）、TIN-边（TIN-edge, P_2）、侧边（side-edge, P_3）、TIN 面（TIN-face, P_4）、侧面（side-face, P_5）和 GTP（P_6）。此外，为了空间操作方便，在 GTP 构模单元中还引入了对角线（diagonals, P_7），其目的是在空间切割与计算时将一个 GTP 切割为 3 个四面体。因此，一般情况下，一个 GTP 包含：3 个上结点、3 个下结点、3 个上 TIN 边、3 个下 TIN

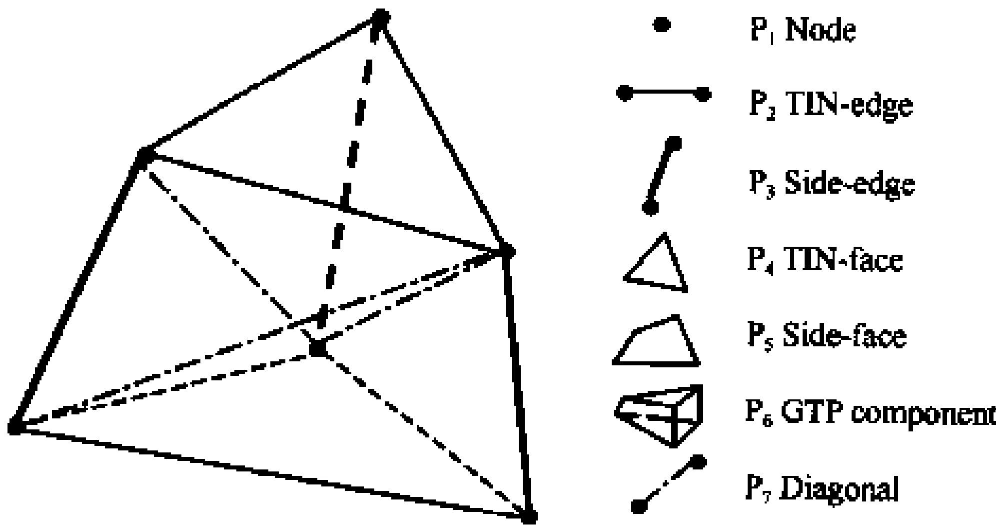


图 4-15 GTP 单元的组成要素

边、1个上TIN面、1个下TIN面、3条侧边、3个侧面和3条对角线。

由于钻孔偏斜，侧面不一定总是平面，或者说每个侧面的4个结点不一定共面。引入对角线的目的，就是为了处理这种情况：对角线将每一侧面分割为两个三角形，并将一个GTP单元切割为3个四面体，从而便于空间操作和空间分析，包括体积计算、空间剖切、包含查询及地质统计分析等。为了减少数据冗余，对角线数据并不在数据库中存储，只是当需要进行空间操作与空间查询时动态生成。

4.3.2 GTP构模特点

GTP构模的特点是：

(1) 基于采样数据：直接利用钻孔采样数据，而无需进行空间内插，即可通过钻孔采样数据以TIN的形式来模拟和表达地层界面的基本空间形态，最大限度地保障3D地学构模精度。

(2) 开放式构模：当有新的钻孔数据或通过物探、化探、测量等手段获得新的地层空间信息时，只需在局部进行修改或扩充，而不需改变整体结构，这样使得GTP的局部细化与动态维护很方便。

(3) 有拓扑描述：以GTP为基本的体元构模单元，便于进行拓扑描述与表达。

(4) 基于TIN的2.5D GMS为其子集：同一地层GTP集合的上下界面为TIN结构，因此基于TIN的2.5D GMS可以看成基于GTP的3D GMS的一个子集。

(5) Pyramid、TEN模型为其退化：基于TIN边退化和TIN面退化，可以由GTP分别导出Pyramid模型和TEN模型，如图4-16所示。当某一条侧边收缩为一个结点时，GTP退化为Pyramid；当某一个TIN收缩为1个结点时，GTP退化为四面体。GTP的这一特点非常适合处理地层尖灭、分叉、断层切削等复杂情况。

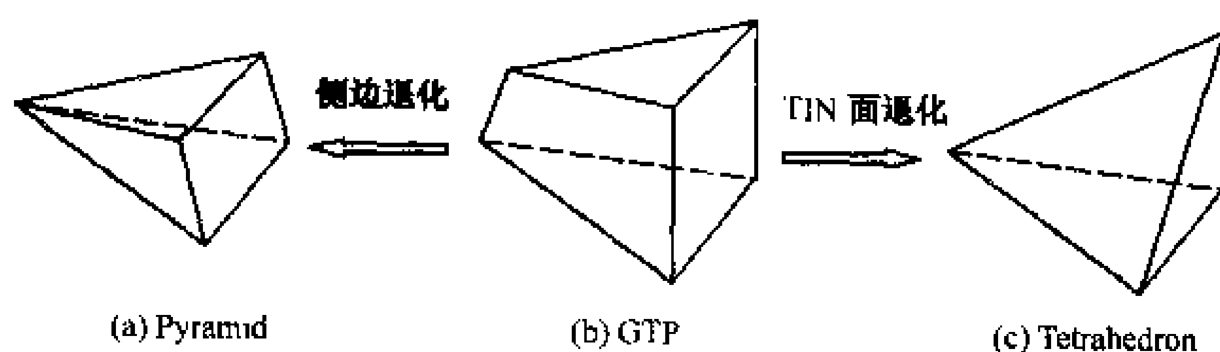


图4-16 GTP的两种退化模式

GTP构模与前述三大类构模方式的区别与联系在于：GTP首先是体元构模方式，它不仅能构建3D地层体和地层TIN面，还可以进行拓扑描述。而单一的面模型只能构建地质体的表面；单一的体模型却很难表达地质体表面形状，且拓扑描述研究少；面-体混合模型则是在一个系统里面同时有两种构模方式，系统存储和操作均不方便。

4.3.3 GTP构模过程

基于GTP的3D地学构模过程如图4-17所示。图中，结点代表由钻孔揭露的地层

界面点，为初始的0D要素；TIN边和侧边分别为连接沿界面或界面之间结点的1D要素；TIN面和侧面则为由TIN边和侧边闭合而成的2D要素；TIN面、侧面封闭形成3D要素——GTP。一组具有相同属性的GTP组合在一起（侧面相邻和/或TIN面相邻），生成1个地质体如地层、矿体、地质结构等。若干地质体组合起来，即形成研究区域的3D复杂地质模型。

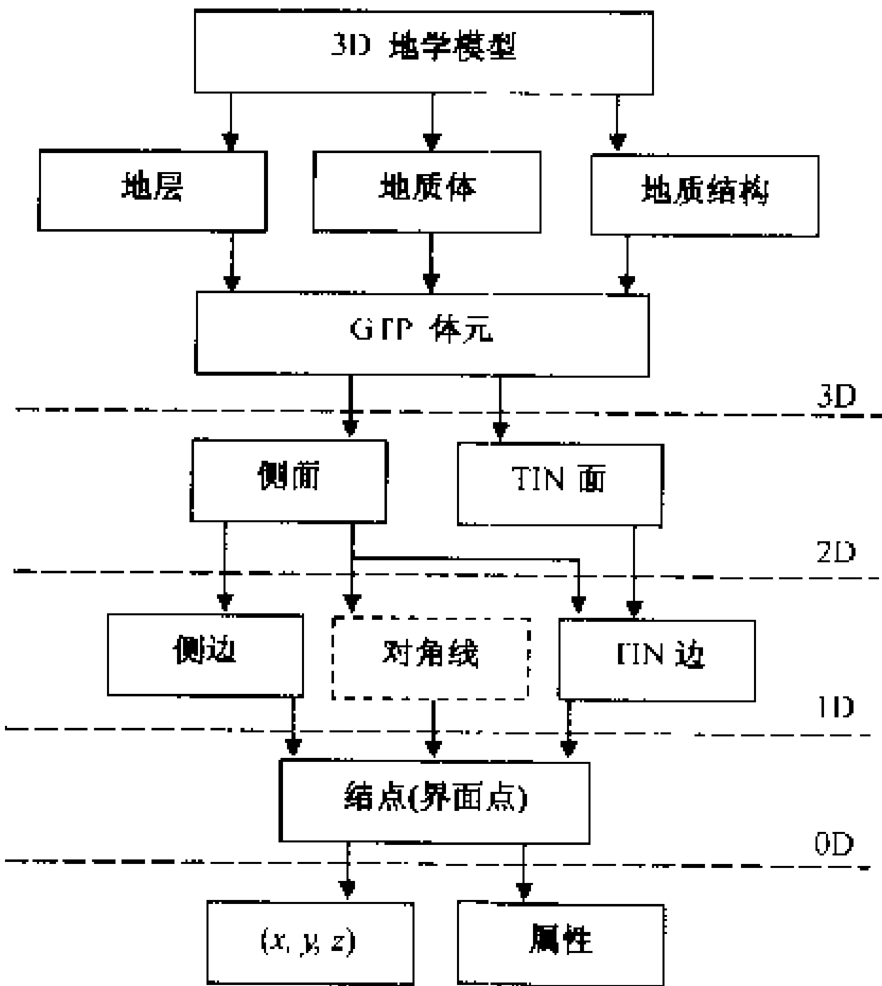


图4-17　基于GTP的3D地学构模过程

4.3.4 GTP数据结构

GTP由结点、TIN边、棱边、TIN面、侧面和GTP体共6个基本元素组成，其数据结构如下：

1. 结点数据结构

```
Type
CNode = Record
  PID: integer;   //ID label of node
  FNode: integer;   //the feature of node, "hard" or "soft"(if the node is a sampling point, it is hard and its position can no longer be move; if the node is a deduce or geo-interpreted point, it is soft and its position can be adjusted if demanded)
  UpBelong, LoBelong: integer;   //ID label of related upper layer and lower layer
  x, y, z: float;   //position
  next: ^long;   //ID-pointer to the next Node in chain table
  TEdge[n]: array of TIN-edge;   //adjacent TIN-edges
```

```
  UpSEdge, LoSEdge:CSide-edge;   // adjacent upper side-edge and lower side-edge
End。
```

2. TIN 边数据结构

```
Type
CTIN-edge = Record
  TEID:integer;   //ID label of TIN-edge
  next:^long;   //ID-pointer to the next TIN-edge in chain table
  Pstart, Pend: CNode;   //start-node and end-node of TIN-edge
  LeftTIN-face, RightTIN-face: CTIN-face;   //left TIN-face and right TIN-face
  UpSideFace, LoSideface:CSide-face;   //upper side-face and lower side-face
  Bcut:Boolean;   //label if clipped
  Pcut: TPoint;   //the clipping point
End。
```

3. 侧边数据结构

```
Type
CSide-edge = Record
  SEID:integer;   //ID label of side-edge
  next:^long;   //the ID-pointer to the next side-edge in chain table
  Pstart, Pend: TNode;   //start-node and end-node of side-edge
  AdjSFace[n]:array of side-face;   //its adjacent side-faces
  Bcut:Boolean;   //label if clipped
  Pcut: TPoint;   //the clipping point
End。
```

4. TIN 面数据结构

```
Type
  CTIN-face = Record
  TINID:Integer;   //ID-label of TIN-face
  Node[3]: TNode;   //its three nodes
  TINEdge[3]:TTIN-edge;   //its three TIN-edges
  AdjacentTIN-face[3]:Array of Integer;   //the ID-label of its three adjacent TIN-
face
  UpAdjGTP, LoAdjGTP:CAdjGTP;   //its upper adjacent and lower adjacent GTP
  Bcut:Boolean;   //label if clipped
  Pcut: TTIN-edge;   //the clipping TIN-edge
End。
```

5. 侧面数据结构

```
Type
CSide-face = Record
  SFaceID:integer;   //the ID-label of side-face
  Node[4]: TNode;   //its four nodes
  TIN-edge[2]:CTIN-Edge;   //its two TIN-edges
  Side-edge[2]:CSide-Edge;   //its two side-edges
  LeftGTP, RightGTP:CAdjGTP;   //its adjacent left GTP and right GTP
  Bcut:Boolean;   //label if clipped
  Pcut: TSide-edge;   //the clipping side-edge
End。
```

6. GTP 数据结构

```
Type
CGTP = Record
  GTPID:Integer;   //the ID-label of GTP
  LayerID: Integer;   //the ID-label of layer belonged
  Node[6]: TNode;   //its six nodes
  UpTIN-face, LoTIN-face: CTIN-face;   //its upper TIN-face and lower TIN-face
  Side-face[3]:CSide-face;   //its three side-faces
  AdjGTP[5]:Array of integer;   //its five face-adjacent GTP components
  Bcut:Boolean;   //label if clipped
  Pcut1: TTIN-face;   //the clipping TIN-face
  Pcut2: TSide-face;   //the clipping side-face
End。
```

§4.4 基于 GTP 的 3D 模型更新

前已述及，GTP 构模的特点之一是“开放式构模”，即当有新的钻孔数据或通过物探、化探、测量等手段获得新的地层空间信息时，只需在局部进行修改或扩充，而不需改变整个体的结构。空间数据局部更新方法与空间数据来源有关。

4.4.1 加入新钻孔的模型更新

当有新的钻孔点数据加入时，首先要判断新加的钻孔点的位置，分 3 种情况：

(1) 新钻孔点在原 TIN 的边界之外：此时只需将新钻孔点追加到离散点数据库的最后，如图 4-18 所示。

(2) 新钻孔点在原 TIN 的边界之内：此时需要判断新点位于哪个三角形内。判断

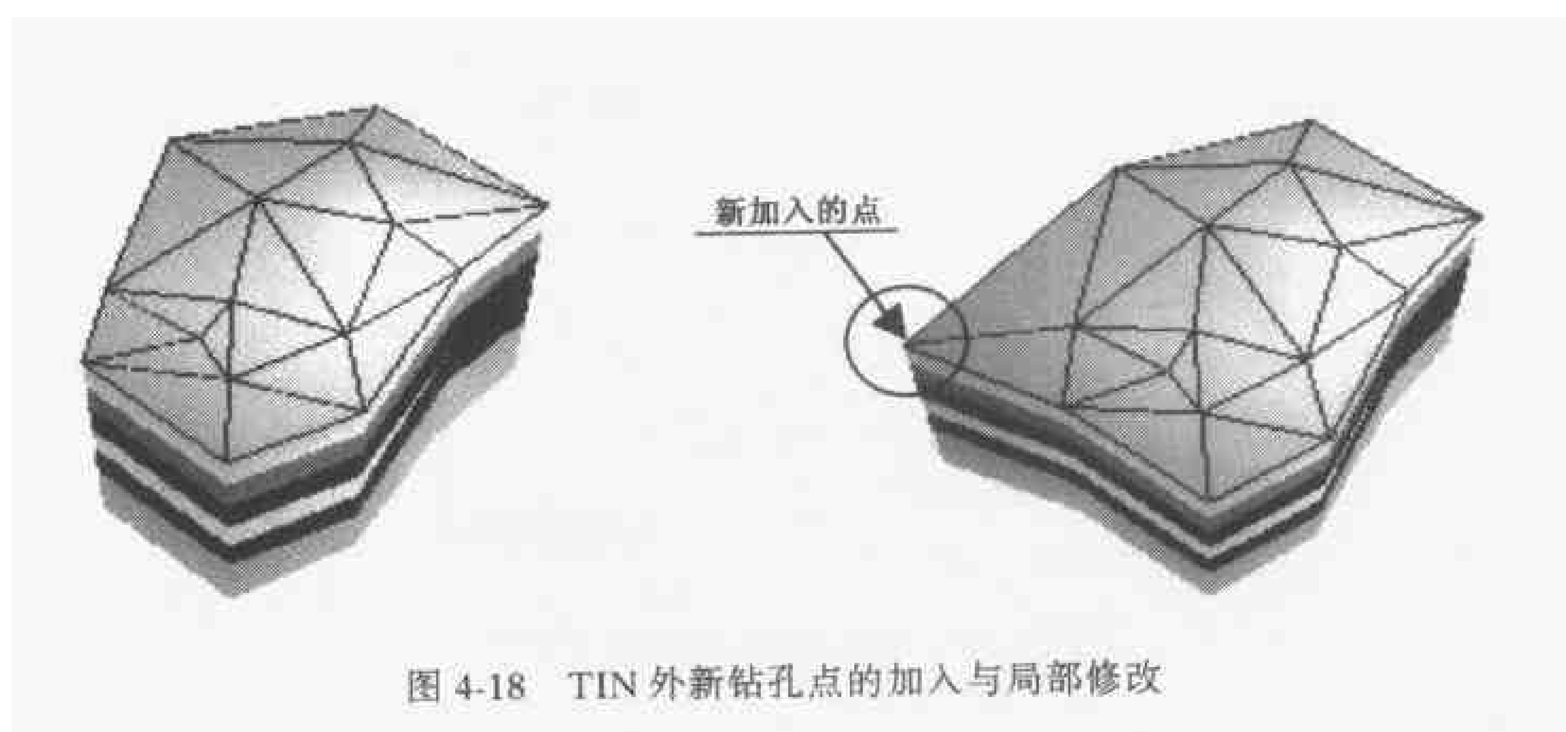
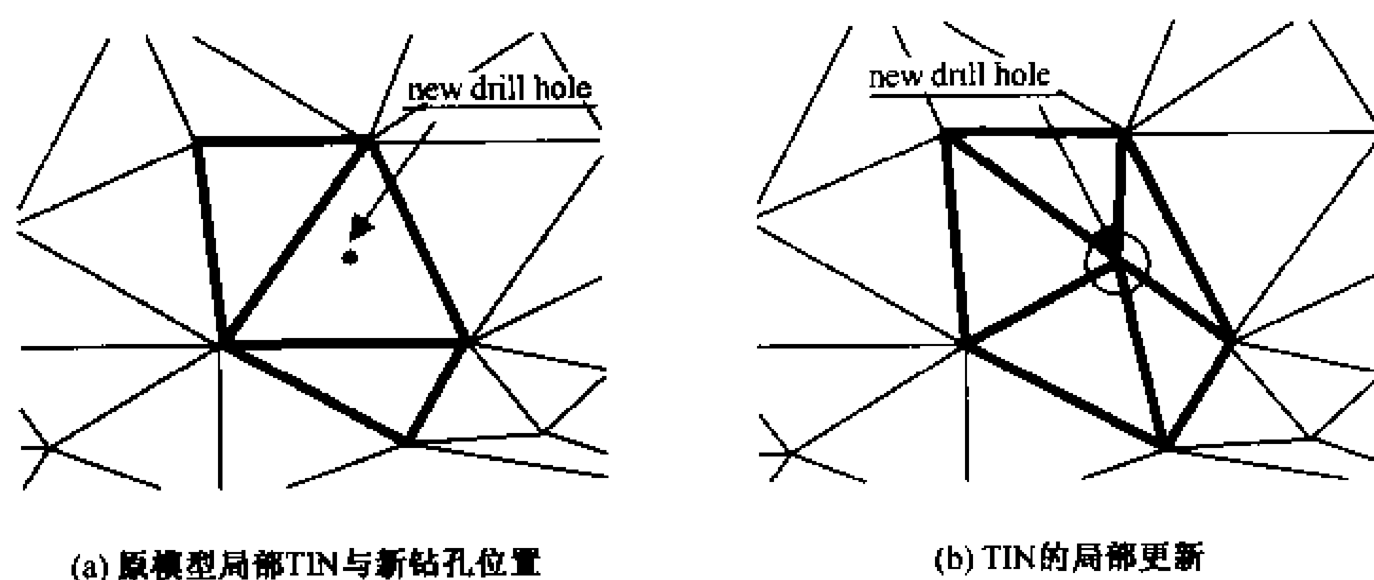


图 4-18　TIN 外新钻孔点的加入与局部修改

点在三角形内可遵循如下法则：a）点到三角形的距离为 0；b）新点和三角形的 3 个顶点组成的 3 个小三角形面积的和等于原来大的三角形面积。判断出新点在哪个三角形之后，将该点插入到所处三角形的 3 个顶点之间即可，并在数据库记录里添加该点的数据，如图 4-19 所示。点的位置判断出来以后，就可以在局部修改 TIN 的生成以及在局部修改 GTP 的生成。



(a) 原模型局部TIN与新钻孔位置　　(b) TIN的局部更新

图 4-19　TIN 内新钻孔点的插入与局部修改

TIN 及 GTP 的局部重构方法为：当判断出新加入的点所在的三角形的位置以后，新的点将原来的一个三角形分成 3 个三角形，然后判断 3 个新的三角形是否符合 Delaunay 法则，若是则保留，若不是则去掉，重新和其他的点构建三角形，之后再判断是否符合 Delaunay 法则。逐渐往外扩散，递归执行这种操作，直到所有新的三角形都符合 Delaunay 法则为止。TIN 的局部重构完成以后，对应可完成 GTP 的局部重构。

(3) 新钻孔点揭露新的地层：当新钻孔点揭露了当前模型没有表达的新地层时（如在巷道里打暗钻并发现新的地层时），需要将新的地层加进去。首先对新地层的平面发育范围与形态进行地质推估，并判断新层与已有模型的立面位置关系；然后在原层的局部范围添加了该新地层，如图 4-20 所示。此时需要在新地层界面上构建 TIN，进而构建新的 GTP，从而达到构建新地层的目的。

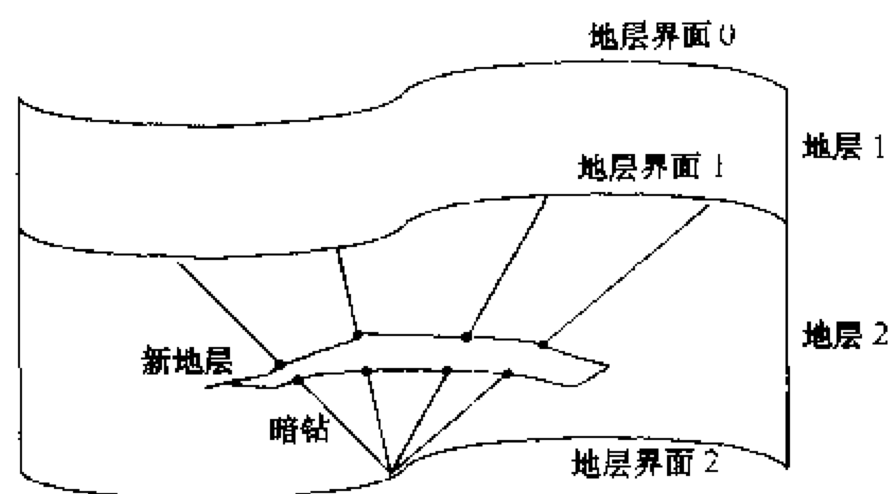


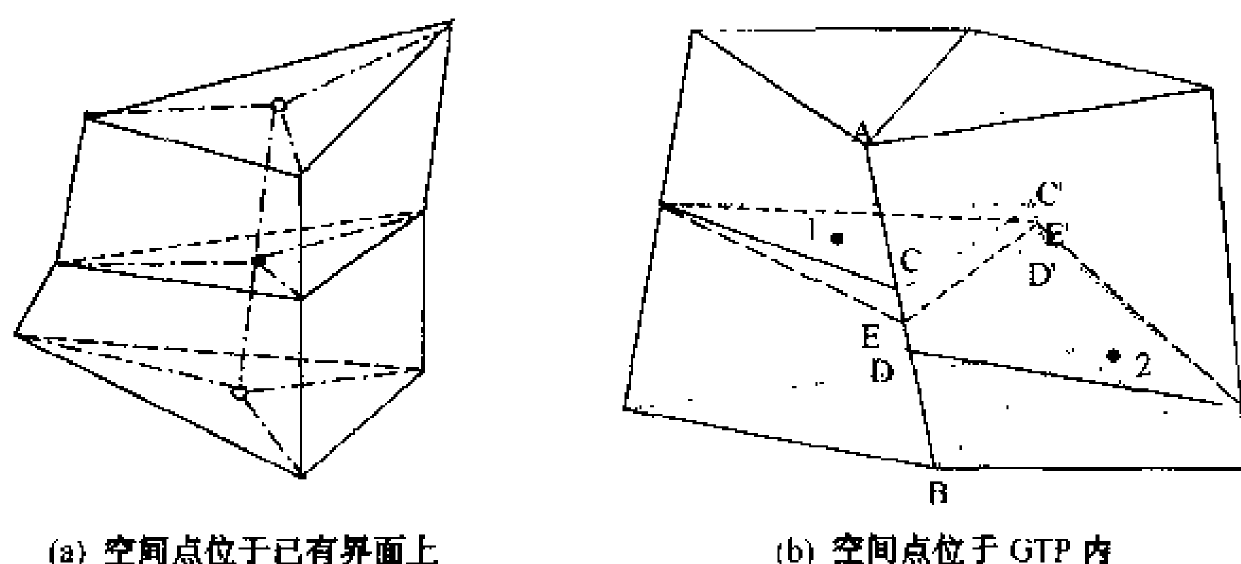
图 4-20　新地层的空间推估示意图

4.4.2　加入新空间点的模型更新

当有新的空间点（此处泛指通过物探、化探、测量等手段获得或开挖揭露的新的地层空间信息）加入时，同样需要首先判断点的位置。也分两种情况：

(1) 空间点位于已有界面上：此时先在该界面上判断点位于哪个三角形内，判断方法同前；其次局部重构 TIN；然后根据相邻钻孔迹线的方向，在上一界面和下一界面上对应地内插点；最后局部重构 TIN 及 GTP，同时进行拓扑修改，如图 4-21a 所示。

(2) 空间点位于 GTP 内：首先根据点的走向、倾向方位角，向该点临近的 3 个钻孔投影，得到 3 个点（相当于由点所在的面向外扩展和钻孔相交，得到 3 个交点）；其次将这 3 个点构成 TIN，和它相邻的上下三角形构成两个 GTP；然后进行 TIN 和 GTP 重构。值得指出的是：当同一层的另外一个空间点按照上述方法也得到一个新的三角形，而这个三角形和刚才的三角形在同一个钻孔上的交点不重合，这时可以在同一钻孔上取这两个交点的中点，让这两个交点强制拟合即可。如图 4-21b 所示，C 点是空间点 1 在钻孔 AB 上的投影点，D 点是空间点 2 在钻孔 AB 上的投影点，显然，C 点和 D 点不重合，C′点和 D′点不重合，此时，可以取 C、D 的中点 E，将 C、D 两点合二为一，从而保持两个新面的连续性。



(a) 空间点位于已有界面上　　(b) 空间点位于 GTP 内

图 4-21　非界面空间点加入后的 GTP 重构原理

§4.5 基于GTP的真3D地学模拟

4.5.1 系统开发与关键技术

利用 VC++ 和 OpenGL 将 GTP 构模原理与方法进行了程序实现，开发了 3D 地学建模系统 GeoMo3D。该系统解决了钻孔偏斜计算、插值计算、Delaunay TIN 的生成、坐标旋转、平移、光照处理、放大缩小、漫游等基本问题。并基于地学数据来源设计了多源集成的地学数据库，其中钻孔数据库由钻孔号、岩层编号、岩层倾角、岩层方位角、岩层名称、岩性描述、岩层厚度、岩层界面坐标等基本字段组成。从地学数据库里直接提取数据，通过 2D 地质解译和可视化交互编辑，运用 GTP 数据结构生成 3D 地质体，在此基础上实现了地质模型的 3D 可视化、3D 拓扑查询、揭层显示、任意剖面切割、篱笆图生成等主要 3D 空间操作功能。

本系统开发的两项关键技术是：钻孔测斜计算与 GTP 切割计算。

1. 钻孔测斜计算

数据预处理的关键是选择测斜公式，计算钻孔与地层交点的倾斜角和方位角，如图 4-22 所示。由于测斜点的（x，y，z）坐标一般不可知，故采用数值积分的方法作为测斜计算公式。具体的思路是：在两个测斜点之间将倾斜角和方位角的增量分成若干段，建立积分公式：$\Delta\theta_l=\int_0^l \frac{\Delta\theta}{L}\mathrm{d}l$ 和 $\Delta\varphi_l=\int_0^l \frac{\Delta\varphi}{L}\mathrm{d}l$（$L$ 表示孔口到测斜点处的迹线长度，l 表示孔口到所求层的迹线长度，$\Delta\theta$ 表示测斜角），然后按照这两个积分公式，计算该段内每一岩层的倾斜角和方位角增量。最后得到的钻孔迹线是经过测斜点的光滑曲线。

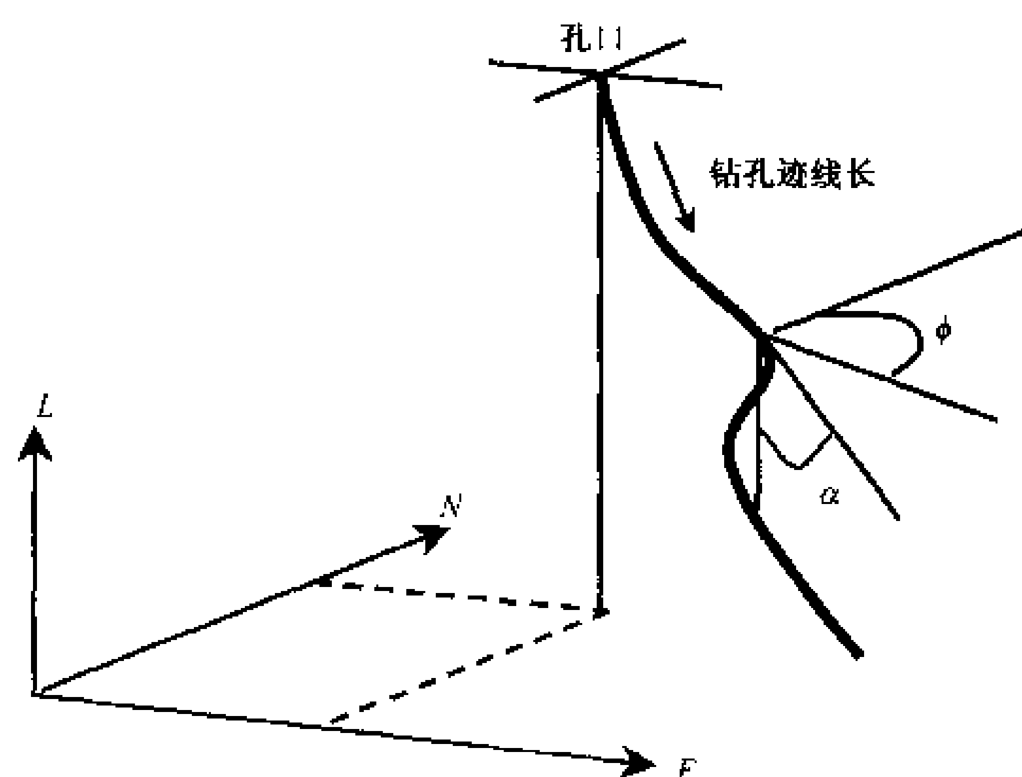


图 4-22 钻孔测斜计算原理
（据 Simon W. H，1994）

2．GTP 切割计算

在 3D 空间中进行任意方向的剖切，可以对地层空间进行任意角度的观察。剖切的关键是如何形成剖面多边形。剖切算法步骤如下：

(1) 预处理：首先对每个 GTP 体元，确定其边构成；然后对每个体元中各边的两个顶点进行判断，确定一个标记符，标记该顶点与剖切平面的相对位置：若该顶点到剖切平面的距离大于零，则记为 1；否则，记为 0。

(2) 交点计算：对于每个 GTP 体元，逐个判断其中的每一条边。如果边的两个顶点标记不同，则该边与剖切平面相交，计算交点的 3D 坐标值。

(3) 剖面多边形的形成：如果一个体元不存在交点，则该体元不对剖面多边形做任何贡献，否则存在交点。提取任意一个交点作为剖面多边形的第一个顶点，依据边-面和面-边之间的拓扑关系，找到与该交点所在边共面且位于该体元中的另一个具有交点的边，将第二个交点作为剖面多边形的下一个顶点。以此类推，直至搜寻回到第一个顶点。这样就形成了体元中的剖面多边形。

图 4-23 所示为 GTP 与剖切平面相截的 12 种典型情形。

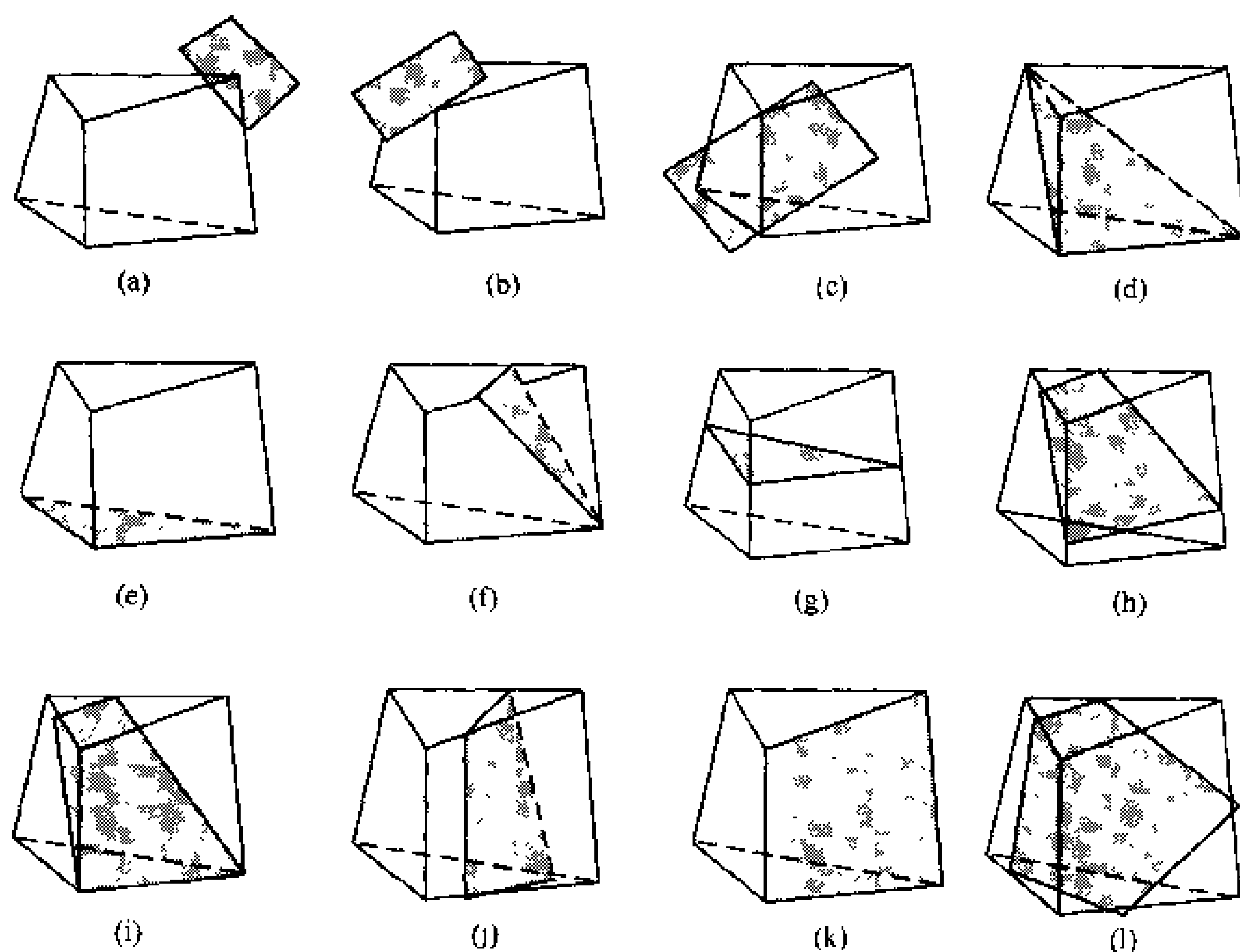


图 4-23　GTP 剖面切割的 12 种典型方式

(a) 与 Node 相切；(b) 与 TIN-edge 相切；(c) 与 Side-edge 相切；(d) 与两对角线相切；(e) 与 TIN-face 重合；(f) 与 Node、TIN-face 及两 Side-face 相切；(g) 与 3 个 Side-faces 相切；(h) 与 1 个 TIN-face、3 个 Side faces 相切、(i) 与 TIN-edge、TIN face 及 2 个 Side-faces 相切；(j) 与 2 个 TIN faces、2 个 Side-faces 相切；(k) 与 Side-face 重合；(l) 与所有面相切

4.5.2 真 3D 地学模拟实例

以开滦（集团）有限责任公司的钻孔资料数据为数据源。经测斜计算处理得到钻孔迹线的光滑曲线如图 4-24 所示。

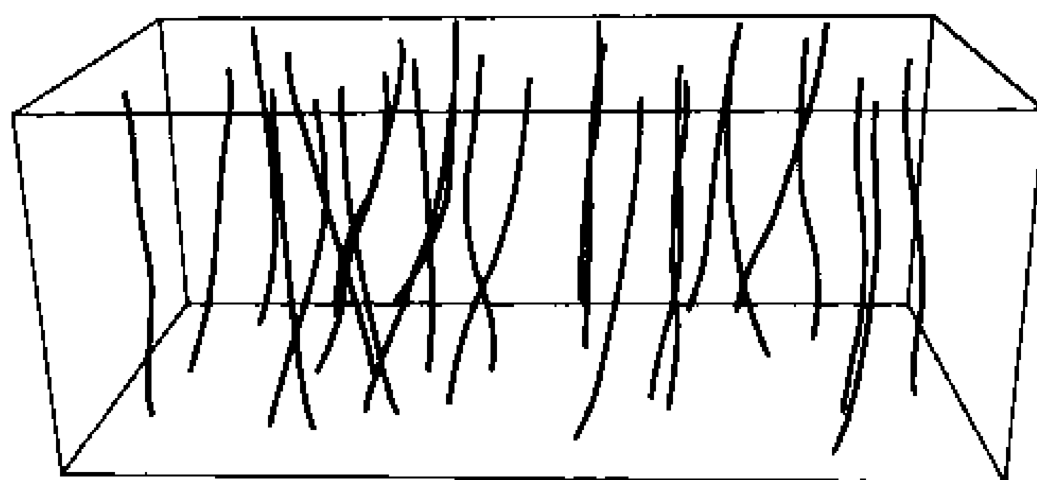


图 4-24 钻孔迹线的 3D 可视化

经地层面 TIN 自动生成、地层系列 GTP 构建、GTP 充填、颜色选择、光照处理等基本过程，得到如图 4-25 所示的区域 3D 地层模型。图 4-26 所示为对所构模型的一种任意剖切结果。

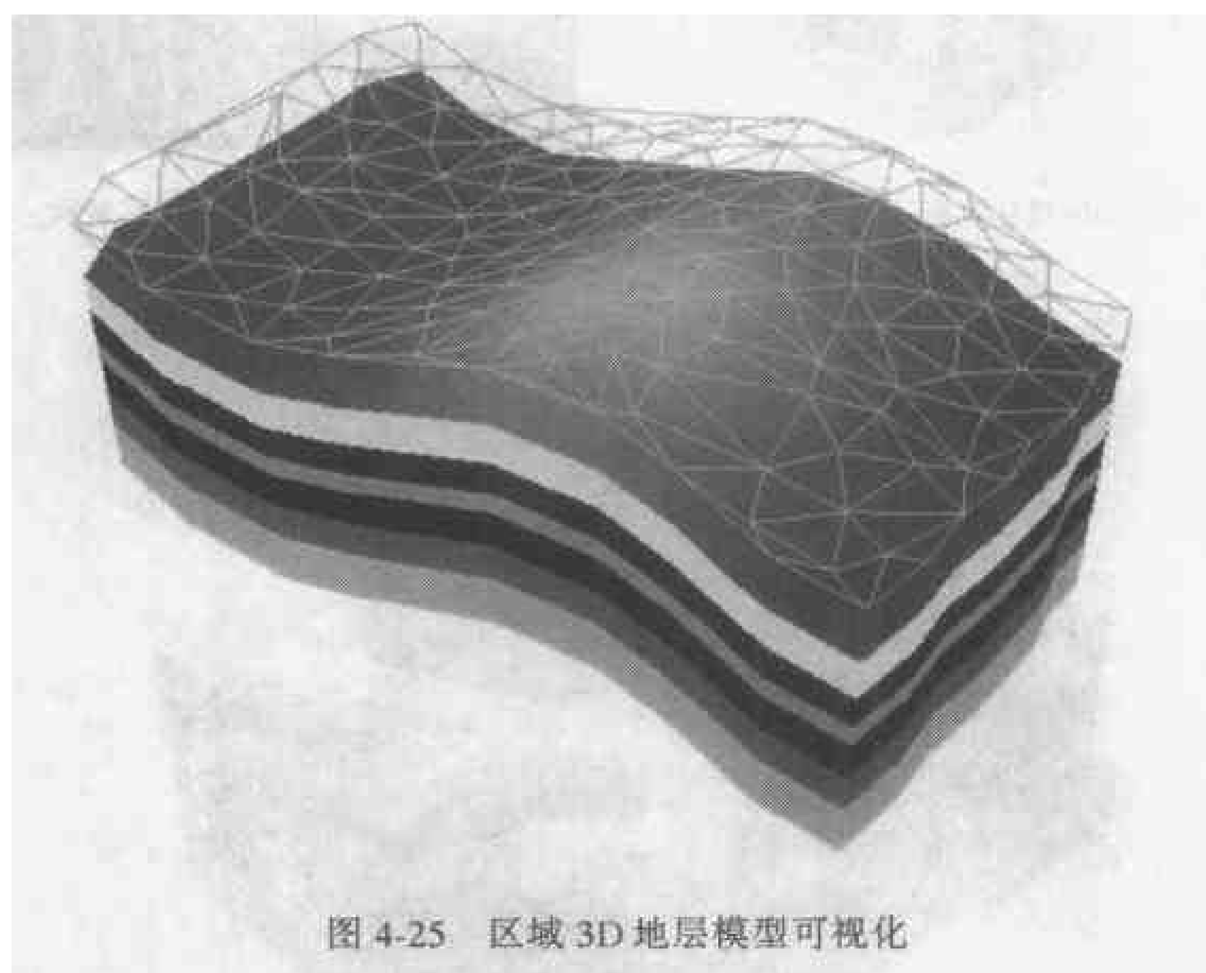
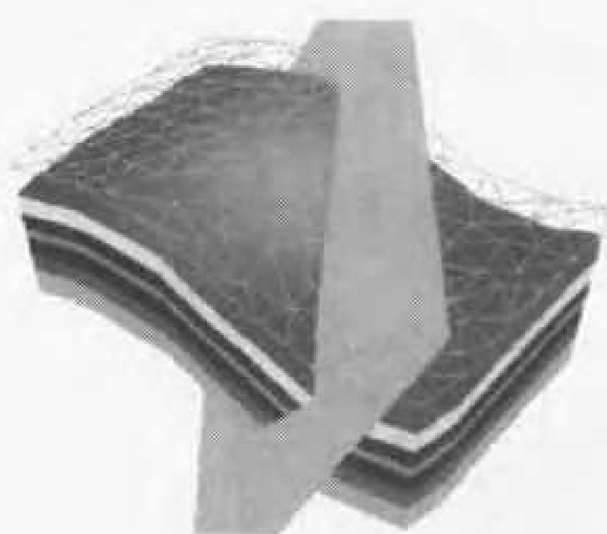


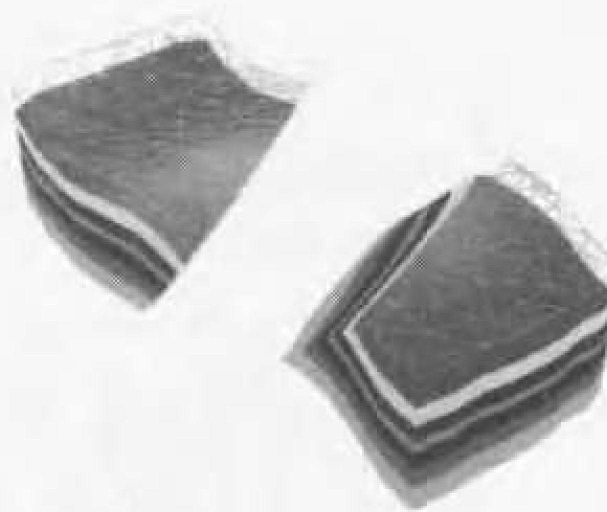
图 4-25 区域 3D 地层模型可视化

将层状地质体进行分层或揭层显示是地学可视化中的一项重要内容。选择性地抽去地层序列中某些地层而保留主要地层，或在剖面切割之后将某一段块的地层序列中某些地层揭走，可以清晰地看到地质体各层情况及其空间组合关系，如图 4-27 所示。

在分层或揭层显示的基础上，再进行多而组合切割形成地质篱笆图，还可以进一步了解地质体内部的精细信息，如图 4-28 所示，这对于开采、开挖设计与评估非常有益。

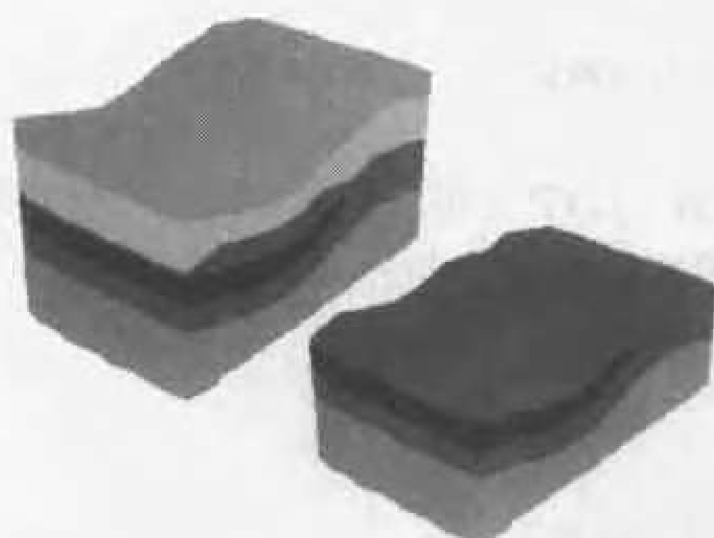


(a) 2D剖面位置

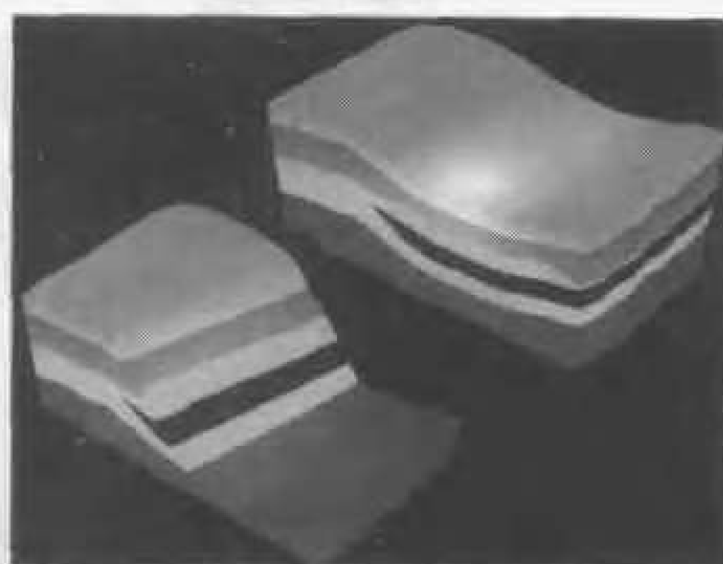


(b) 切割并分离

图 4-26　基于 GTP 的 3D 地层模型的剖面切割可视化



(a) 3D地层分层显示



(b) 3D地层切割后揭层显示

图 4-27　基于 GTP 的 3D 地层的分层、揭层与切割可视化

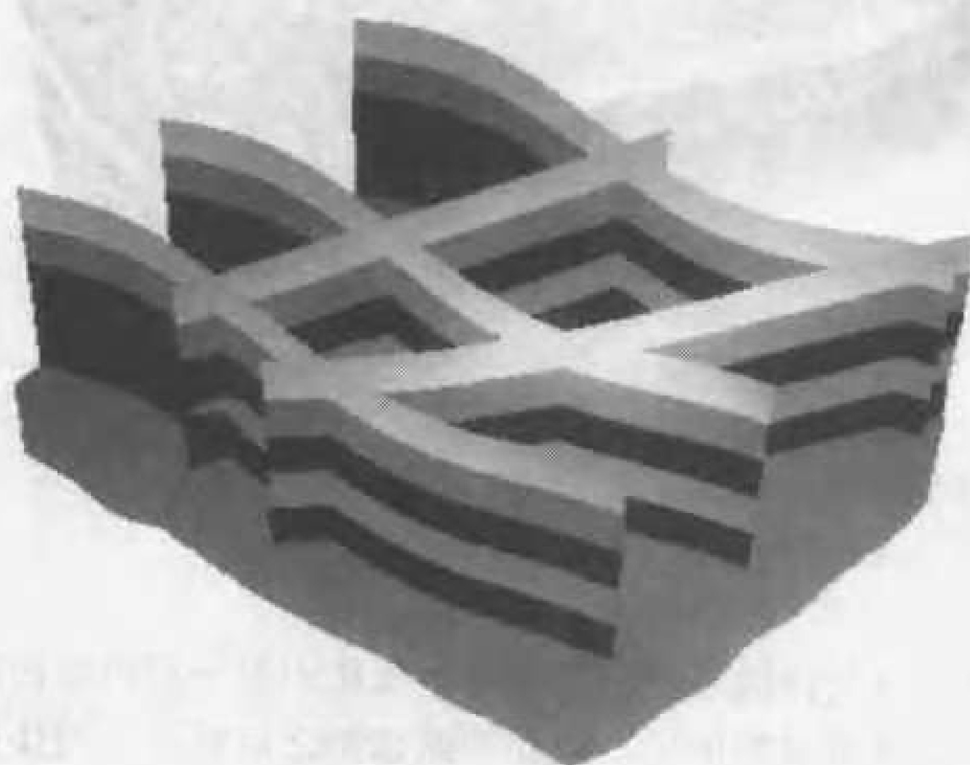


图 4-28　基于 GTP 的 3D 地层组合切割与篱笆图可视化

3D GIS 以地球表面及以上为研究对象，而 3D GMS 以地球表面及以下为研究对象，两者结合起来即为地球整体。3D GIS 与 3D GMS 在发展道路是并行的且互相影响，3D GIS 已在吸收 3D GMS 的养分来加强其 3D 空间构模研究，而 3D GMS 也在吸收 3D GIS 的养分来促进其 3D 拓扑描述研究。可以预见，在不远的将来两者将殊途同归，最后为全要素真 3D 地球信息系统的实现提供理论、方法与技术基础。届时，可以在一个统一的时空框架下对地球以上、地球表面及地球内部进行整体的真 3D 可视化表达、管理、更新、查询、分析与操纵。目前，围绕数字城市(Shi W. Z.,1996, 2000,2003;孙敏等,2000)、数字矿山(吴立新,2000;吴立新等,2000,2003)和数字海洋(Christopher C. M.,2000)进行地球表面上下空间统一构模方面的研究已经展开，这是迈向未来的第一步，任重而道远。

参 考 文 献

边馥苓,傅仲良,胡自锋,2000. 面向目标的栅格矢量一体化 3D 数据模型. 武汉测绘科技大学学报,25(4): 294~298

陈 健,2001. 3D 地层空间信息系统结合 3D 有限元的研究与实现(博士学位论文). 中科院武汉岩土所

戴吾蛟,邹铮嵘,2001. 基于体素的 3DGIS 数据模型的研究. 矿山测量,(1): 20~22

龚健雅,夏宗国,1997. 矢量与栅格集成的 3D 数据模型. 武汉测绘科技大学学报,22(1): 7~15

韩国建,郭达志,金学林,1992. 矿体信息的八叉树存储和检索技术. 测绘学报,21(1): 13~17

赫尔丁 S. W.,1989. 3D 矿床的计算机构模方法. 国外金属矿山,(2): 95~96

侯恩科,吴立新,李建民等,2002. 3D 地学模拟与数值模拟的耦合方法研究. 煤炭学报,27(4): 388~392

侯恩科,吴立新,2002. 面向地质建模的 3D 体元拓扑数据模型研究. 武汉大学学报(信息科学版),27(5): 467~472

胡金星,吴立新,杨可明等,1999. 3D 地学模拟体视化技术应用研究. 煤炭学报,24(4): 345~349

惠勒 A. J.,斯托克斯 P. C.,1989. 块段模型和线框模型在地下采矿中的应用. 国外金属矿山,(2):98~101

李德仁,李清泉,1997. 一种 3DGIS 混合数据结构研究. 测绘学报,26(2): 128~133

李清泉,1998. 基于混合结构的 3DGIS 数据模型与空间分析研究(博士学位论文). 武汉测绘科技大学

李清泉,李德仁,1998. 3D 空间数据模型集成的概念框架研究. 测绘学报,27(4): 325~330

李青元,林宗坚,李成民,2000. 真 3DGIS 技术研究的现状与发展. 测绘科学,25(2): 47~51

齐安文,吴立新,侯恩科等,2001. 3D 地学模拟述评及其矿山应用关键问题. 中国矿业,10(5): 61~64

齐安文,吴立新,李 冰等,2002. 一种新的 3D 地学空间构模方法:类三棱柱法. 煤炭学报,27(2): 158~163

孙 敏,陈 军,2000. 基于几何元素的 3D 景观实体建模研究. 武汉测绘科技大学学报,25(3): 233~237

孙 敏,陈 军,张学庄,2000. 基于表面剖分的 3DCM 空间数据模型研究. 测绘学报,29(3): 257~265

王海峰,2000. 3D 空间信息的数据生成算法研究和实现[硕士学位论文]. 北京大学

吴健生,2001. 地质体三维可视化及空间数据探索[博士学位论文]. 中国科学院地理科学与资源研究所

吴立新,2000. 数字地球、数字中国与数字矿山. 矿山测量,(1): 6~9

吴立新,殷作如,邓智毅等,2000. 论 21 世纪的矿山——数字矿山. 煤炭学报,25(4): 337~342

吴立新,张瑞新等,2002. 3D 地学模拟与虚拟矿山系统. 测绘学报,31(1): 28~33

吴立新,史文中,Christopher G. M.,2003a. 3D GIS 与 3D GMS 中的空间构模技术. 地理与地理信息科学,19(1): 5~11

吴立新,殷作如,钟亚平,2003b. 再论数字矿山:特征,框架与关键技术. 煤炭学报,28(1): 1~7

肖乐斌,龚健华,谢传节,1998. 线性四叉树和线性八叉树邻域寻找的一种新算法. 测绘学报,27(3):195~203

许 斌,张 森,历万庆,1994. 从序列切片重构 3D 对象的新方法. 计算机学报,17(1): 64~71

张 煜,白世伟,2001. 一种基于三棱柱体体元的 3D 地层建模方法及应用. 中国图像图形学报,6(3): 285~290

赵树贤,1999. 煤矿床可视化构模技术[博士学位论文]. 中国矿业大学(北京校区)

朱小弟,李青元,曹代勇,2001. 基于 OpenGL 的切片合成法及其在 3D 地质模型可视化中的应用. 测绘科学,26(1): 30~32

朱小弟,2001. TITIAN 三维建模软件. 北京东方泰坦科技有限公司产品资料

阿列尼切夫 B. M.,弗拉基米罗夫 A. N.,1995. 菱镁矿股份公司露天采矿工程计划编制自动化. 国外金属矿山,(12):70 -74

Arnaud D. L. and Bernard C., 1999, 3D topological modeling and visualization for 3D GIS. Computers & Graphics, 23: 469 ~478

Breunig M., 1999. An approach to the integration of spatial data and systems for a 3D geo information system. Computers & Geosciences, 25(1): 35 ~48

Breunig M, et al., 1999. First steps towards an interoperable GIS -an example from Southern Lower Saxony. Phys. Chem. Earth(A), 24(3): 179 ~189

Christopher G. M. and Maydell U. M., 1978. Triangulation and spatial ordering in computer cartography. Proc. 3rd Canadian Cartographic Association Annual Meeting. Canada, 69 ~ 81

Christopher G. M., 2000. Chapter 4: An Algorithmic Approach to Marine GIS. In Marine and Coastal Geographical Information Systems(Eds, Wright, D. and Bartlett, D.)37 ~52

Christopher B. J., 1989. Data structure for 3D spatial information system in geology. Int. J. GIS, 3(1): 15 ~31

David W. P. et al., 2000. Data fusion modeling for groundwater system. Journal of Contaminant Hydrology. 42(2 ~4), 303 ~335

Ferenc S., 1994. The GIS concept and 3-dimensional modeling. Computers, Environment and Urban Systems, 18(2): 111 ~121

Fisher T. and Wales R.Q., 1991. 3-D solid modeling of geological objects using non-uniform rational B-splines (NURBS). In Three Dimensional Modelling with Geoscientific Information Systems, edited by Turner A.K. (Kluwer, Dordrecht), pp. 85 ~105

Gabriel C. et al., 2001. 3D volumetric modelling of Cadomian terranes (Northern Brittany France): an automatic method using Voronoi diagrams. Tectonophysics. 331: 181 ~196

Gong, J. Y., Cheng P. G., Liu R., Li D. J. and Liu S. H., 2002. Study on 3D modeling and visualization in geological exploration engineering. In Proc. ISPRS Commission II Sym. on integrated system for spatial data production, custodian and decision support, 133 ~138

Homer H. and Thomas S., 1998. A survey of construction and manipulation of Octree. CVGIP, 43: 409 ~431

Joe B., 1991. Construction of 3D delaunay triangulations using local transformations. Computer Aided Geometric Design, 8: 123 ~142

Johnson K. R, 1996. Geological modelling - the way ahead (concept modelling for exploration and mine planning). Int. J. Rock Mech. Min. Sci. & Geomechanics Abstract, 33(5): 219A

Jurgen D., 2000. An object-oriented approach for integrating 3D visualization systems and GIS. Computers & Geosciences, 26(1): 67 ~76

Kevin P. and Davide D., 2002. Complex reservoir geometries in a deep water clastic sequence, Gryphon Field, UKCS: injection structures, geological modelling and reservoir simulation. Marine and Petroleum Geolog, 19(2): 161 ~179

Li R. X., 1994. Data structure and application issues in 3D geographic information systems. Geomatica, 48(3): 209 ~224

Mallet J. L., 1997. Discrete modeling for natural objects. Mathematical Geology, 29(2): 199 ~219

Mei Po K., 2000. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. Transportation Research Part C: Emerging Technologies, 8: 185 ~203

Monika R. and Günther G., 1997. GIS datasets for 3D Urban planning. Computers, Environment and Urban Systems, 21(2): 159 ~173

Mucke E. P., 1998. A robust implementation for 3D delaunay triangulations. Int. J. Computational Geometry and Applications, 8(2): 255 ~276

Nicolas S. and Renato P., 1991. Delaunay triangulation of arbitrarily shaped planar domains. Computer Aided Geometric Design, 8: 421 ~437

Pilout M., Tempfli K. and Molenaar M., 1994. A tetrahedron-based 3D Vector Data Model for Geoinformation. In Advanced Geographic Data Modelling, (Ed., Molenaar M). Sylvia De Hoop: Geodesy Press. pp: 129～140

Rychkun E., 1996. 3D geostatistical modeling applied to radioactive waste site. Int. J. Rock Mech. Min. Sci. & Geomechanics Abstract, 33(8): 366A

Shi W. Z., 1996. A hybrid model for 3D GIS. Geoinformatics, (1): 400～409

Shi W. Z., 2000. Development of a hybrid model for three-dimensional GIS. Geo Spatial Information Science, 3(2): 6～12

Shi W. Z., Yang B. S. and Li Q. Q. 2003. An object-oriented data model for complex objects in three-dimensional geographical information systems. IJGIS (accepted)

Simon W. H., 2000. *Practical Geostatistics*. *Springer-Verlag*

Tipper J., 1977. A method and FORTRAN program for the computerized reconstruction of three-dimensional objects from serial sections. Computers & Geosciences, 3(4): 579～599

Tsai V. J. D., 1993. Delaunay triangulation in TIN creation: an overview and a linear-time algorithm. Int. J. of Geographical Information Systems, 7(6): 501～524

Tsai V. J. D. and Alan P., 1993. delaunay tetrahedral data modeling for 3D GIS application. In Proceedings of GIS/LIS'93, (Minneapolis, ACSM-ASPRS-URISA-AM/FM), pp. 671～680

Turner A. K., 1991. Three dimension GIS. Geobyte, 5(1): 31～32

Turner A. K., 1998. 2D, 2.5D, 3D以及4D之间的差异. 测绘通报, (2): 44～45

Victor J. D., 1993. Delaunay triangulation in TIN creation: A overview and a linear-time algorithm. Int. J GIS, 7(6): 501～524

Victor J. D. and Alan P., 1993. Delaunay tetrahedral data Modeling for 3D GIS application. Proc. GIS/LIS'93, 671～678

Simon W. H., 1994. 3D Geoscience Modeling: Computer Techniques for Geological Characterization. Springer-Verlag

Wu L. X, Hou E. K. and Tang C. A., 1999. Geological data organization for FEM based on 3D geosciences modeling. In Proc. 3rd Int. Symposium on DMGIS. Bankou, 323～325

Wu L. X., 2002. A generalized tri-prism model for 3D geosciences modeling and topological description. In Advanced Workshop on Spatial Information Technology. Hongkong, PolyU.

Wu L. X., 2003. Topological relations included generalized tri-prism (GTP) model for 3D geosciences modeling. Computer and Geosciences (accepted)

第五章　空间关系

空间关系是指地理实体之间存在的与空间特性有关的关系，如度量关系、方向关系、顺序关系、拓扑关系、相似关系、相关关系等，是刻划数据组织、查询、分析和推理的基础。空间关系的描述与表达，是GIS区别于CAD等计算机图形处理系统的主要标志。空间关系的研究，直接影响着GIS系统的设计、开发与应用，受到国际GIS及相关学术界的高度重视。总体而言，目前对空间关系的研究主要集中在空间关系的语义、描述、表达、分析与应用等方面。

§5.1　空间关系表达与描述

5.1.1　空间关系概念与研究进展

早期，人们一般认为GIS空间关系主要分为顺序关系（其中包括方向关系）、度量关系和拓扑关系（Pullar D. et al., 1988; Egenhofer M. et al., 1990; Worboys M. A., 1992）。所谓顺序关系，是指空间目标在空间中的某种排序，如前后、左右、上下、东西南北等。所谓度量关系，是指用某种度量空间的度量尺度来描述空间目标之间的关系，如距离等。所谓拓扑空间关系，是指拓扑变换下的拓扑不变量，如空间目标的邻接与关联关系等。上述3种空间关系表达空间数据之间的一种约束，约束的强弱顺序依次为度量关系、顺序关系和拓扑关系。度量关系属于定量关系，顺序与拓扑关系则属于定性关系。总之，空间关系本身包含的内容很广，可以是由空间实体的几何特性引起，可以是由空间实体的非几何特性所导出，也可以是由空间物体的几何特性和非几何特性共同引起（郭仁忠，2000）。

随着GIS空间关系研究不断深入，人们发现空间实体之间还存在着其他多种空间关系，如相邻（adjacency）关系（Christopher C. M., 1992）、相离（disjoint）关系（Florence J. et al., 1996）、模糊反映空间实体邻近的靠近关系（close, near to）以及反映空间目标运动状态的穿越与进入关系等（Egenhofer M. et al., 1990）。此外，华人学者在空间关系研究方面也多有建树，如胡勇等（1997）定义并研究了最邻近和次邻近关系；赵仁亮等（1998）指出空间存在 k 阶空间邻近关系；Ai T. H.（1998）研究了空间实体之间的不确定性空间关系等；Shi W. Z. 和 Guo W.（1999）将目标之间的关系从确定引入不确定之描述。

近年，时空GIS成为一个热门的研究领域，带动了时空拓扑关系的研究。Egenhofer等（1992, 1995）和Szmurlo M. 等（1998）对空间拓扑关系的渐变规律、时空概念等做了探讨，给出了反映拓扑关系时空变化的最邻近拓扑邻接图（closest topological relationship graph）；舒红等（1997）在总结现有时空拓扑关系研究成果的基础上，研究了基于点集理论的形式化拓扑描述方法。

5.1.2 空间关系描述方法

空间关系描述的基本任务是以数学或逻辑的方法区分不同空间关系，并给出形式化描述，为构造空间查询语言和开发空间分析工具提供基础。目前，空间关系描述方法可以分为三类，即交叉方法（intersection-based model）、交互方法（interaction-based method）和基于 Voronoi 图的混合方法（Voronoi-based method）。

1．交叉方法（intersection-based model）

该法的基本思想是将空间实体分解为若干部分，通过比较两个实体的各组成部分之间的交集去判定其空间关系。Egenhofer M. J. 等（1990）基于点集拓扑理论建立的 9－交模型（9-intersection model，又称 9 元组模型，简称 9I 模型）为其代表。9－交模型将空间目标分为边界（boundary）、内部（interior）和外部（exterior）三部分，通过比较空间目标 A 的边界（∂A）、内部（A°）、外部（A ）和空间目标 B 的边界（∂B）、内部（B°）、外部（B^{-}）之间的交集的内容（空 $\varnothing$ 或非空 $\neg\varnothing$）、维数（dimension）、分块（number of seperations），来分析确定两目标之间的空间关系。

2．交互方法（interaction-based model）

该法的基本思想是直接运用空间目标整体而非部分，来定义和区分空间目标之间的空间关系。代表性的工作为 Randell D. 等（1992）和 Cui 等（1993）基于区域连结的定义并运用逻辑演算法描述的空间区域之间的 8 种空间逻辑关系，并发展了基于空间逻辑的推理机制。在此基础上，Vieu L.A.（1993）进一步研究提出了具有完整语义描述的空间关系与空间推理的形式化框架。这种方法需要预先假定目标之间的可能关系，然后进行惟一性描述，缺点是难以保证空间关系的完备性，不适合动态变化情况。

3．Voronoi 图方法（Voronoi-based model）

陈军等（1997）提出，可以用空间目标的 Voronoi 区域（Voronoi region）来定量空间目标的外部，进而对 Egenhofer M. J. 等（1990）的 9－交模型进行了改进，建立了一种基于 Voronoi 图的新 9－交模型，简称 V9I 模型。该模型既考虑了空间目标的内部和边界，同时又以 Voronoi 区域整体作为空间目标的外部，因而集成了交叉方法和交互方法的优点，是对 9I 模型的改良和发展，克服了 9I 模型无法区分相离关系、难以计算空间目标的补等缺点。

5.1.3 空间关系表达方法

空间关系表达的基本任务是以某种方式显式地存储和组织空间目标之间的空间关系，并构建相应的存取与检索方法。目前研究提出的空间关系表达方法主要有关系表法、二维字符串法、Voronoi 图法、偏序法等。

1. 关系表法

关系表法是早期发展的一种空间关系表达方法，即用关系表来显式地表达和记录结点、弧段与面域之间的拓扑邻接、拓扑关联等关系。如 Arc/Info 等许多商业化软件，仍然采用关系表表达方法。关系表法的局限在于仅能表达面域-弧段、弧段-结点、结点-弧段、弧段-面域之间的简单拓扑关系，而对于其他拓扑关系，则未能顾及，更无法表达复杂目标之间拓扑关系。

2. 二维字符串法

该法的基本思想是采用符号投影的方法，将二维空间目标的边界分别沿 x 轴、y 轴作正射投影，并分别生成有顺序关系的字符串，藉以表达和判断目标间的空间关系。显然，二维字符串法的实质是用一维的方法来解决高维的问题，可以较好的表达空间方向关系，但难以保证空间关系表达的可靠性与完备性。为了弥补这方面的局限，Chang S. K. 等（1989）用目标的最小边界矩形来代替空间目标，并通过采用切割机制和增加空间关系算子，提出了一种改进的二维字符串法，即 2D G-string 法，使之能更好地描述空间目标之间的顺序关系。此后，Lee S.Y. 等（1990）提出了 2D C-string 法，孙玉国（1993）提出了 2D T-string 法，可以更好地表达与操作顺序关系、包围与半包围关系。

3. 偏序法

使用偏序集表达空间关系的实质是将空间目标划分为 3 类基本基元，即 0-simple、1-simple、2-simple，基元之间的构成关系就形成了偏序表达（Kainz W. et al., 1993）。偏序表达方法虽然数据结构较复杂，但层次性强，而且偏序结构中包含了包含关系，对空间实体之间的包含、相交、邻近、相接等空间关系较易获得。陈宜金（1998）针对矿井导线网的单纯复形特点，将其表达为偏序关系集和格，并定义了导线网的包含、相邻、相接等空间关系及其交、并等空间操作。

4. Voronoi 图法

以上空间关系的表达方法一般要求空间目标在空间上相连和相接，难以表达空间相邻而不相接的侧向相邻（lateral spatial adjacency）关系，如街道与街边商店、建筑物 A 与建筑物 B、水池与池边树木等均为相邻而不相接。若将包含 N 个空间目标的区域划分为由 N 个 Voronoi 多边形组成的 Voronoi 镶嵌（tesselation），则可以清晰地表达空间目标的侧向邻近关系。Christopher G. M. 等（1995）、Yang W. P. 等（1996）、陈军等（1997）、李成名等（1998）分别围绕 Voronoi 图法的不同方面进行了研究，有效地促进了该法的研究与发展。Voronoi 图法的优点是便于空间数据的更新与动态分析，即空间目标的修改、插入、移动等均为局部操作，空间实体（如车辆）行进过程中可以动态地确定其相邻目标与障碍物。此外，Christopher G. M.（1997）指出，根据 Voronoi 图所表达的空间邻近关系可以推导或判断出其他空间关系，可以尝试在空间数据建模时并不存储拓扑关系，而在 Voronoi 图中动态建立。

5．其他方法

近年，随着GIS的面向对象数据模型的理论研究与技术发展，基于对象的拓扑关系表达逐渐成为一项重要的研究议题。Chung M. C. 等（1995）在Winged-edge拓扑结构的基础上，运用面向对象技术实现了拓扑关系的表达，提供了图形基元之间的邻近性、连接性和定向性信息。毛建华等（2000）基于对Clarke空间理论的分析，提出了一个空间关系的符号表示及其推理的符号集合，指出空间关系的符号表示集合可望与知识的面向对象表示方法结合起来，进而构建空间定性推理框架，支持GIS的空间智能决策。

此外，刘亚彬等（2001）研究了空间对象拓扑关系的推理方法，并将模糊技术与空间对象之间的拓扑推理算法结合起来，以方便处理GIS中的模糊性和不确定性。刘文飞等（2001）顾及空间数据的动态性和不确定性，研究了模糊动态空间关系的表达与识别途径。

§5.2 空间方向关系

指南针的发明为人类提供了粗略定向的手段，经纬仪和陀螺仪的发明则为人类提供了精确定向的技术。空间方向是人类描述和表达客观世界所必须面对的基本空间问题之一，其研究与应用也是GIS的一项主要任务。早在我国魏晋时期，制图学家裴秀在其制图六体（分率、准望、道里、高下、方邪、迂直）中，就把空间方向关系（准望）作为地图制图的数学基础之一进行了研究。

5.2.1 空间方向与空间方向关系

空间方向关系基于空间方向却区别于空间方向。空间方向是定量的，不受距离和空间分辨率的影响；而空间方向关系则兼备定量和定性特征，受目标形状、距离、空间分辨率和可视区域的影响。

1．空间方向描述

空间方向的描述有定性描述和定量描述两种类型。

定性描述是用若干主方向（Ordinal directions）概略地描述空间方向。主方向之间的夹角 θ 相等，每个主方向代表以该主方向为对称线向其左右分别偏转 $\theta/2$ 的方向区间，如图5-1所示。常用的有四方向描述法、八方向描述法和十六方向描述法。四方向描述是把空间方向分为北（N）、东（E）、南（S）、西（W）4个主方向，主方向之间相差90°；八方向描述则是在四方向的基础上，进一步分出东北（NE）、东南（SE）、西南（SW）和西北（NW）4个方向，每个方向之间相差45°；十六方向描述则是在八方向描述的基础上，进一步分出东北北（NNE）、东北东（ENE）、东南东（ESE）、东南南（SSE）、西南南（SSW）、西南西（WSW）、西北西（WNW）和西北北（NNW）8个方向，每个方向之间相差22.5°，如图5-2所示。

定量描述则是用方位角、象限角等比率量标（Ratio），来精确地给出空间目标之间的方向。其中，方位角是从正北方向起，顺时针旋转到某一位置时所经过的角度，取值范围为0～360°；其4个方位象限按方位角大小依次排列，正好与笛卡尔平面直角坐标系的象限排列顺序相反，如图5-3所示，因而三角函数可以不变地使用。方位角是有方向性的，即与两点A、B的顺序有关，AB的方位角和BA的方位角相差180°。所谓象限角，是指某一射线的方向和直角坐标系的坐标轴所夹的小于等于45°的角度，如图5-4所示。象限角和方位角是等价的，可以相互转换。

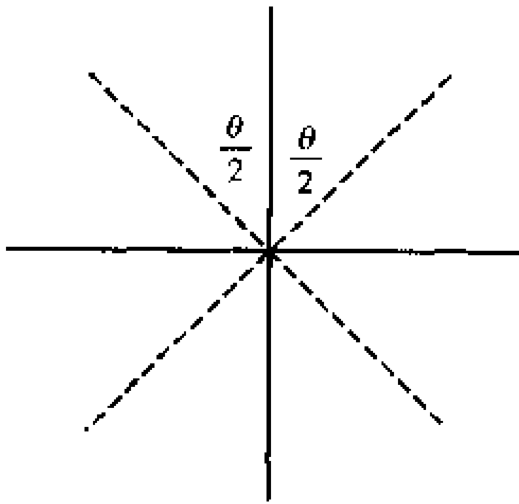


图5-1　主方向描述原理

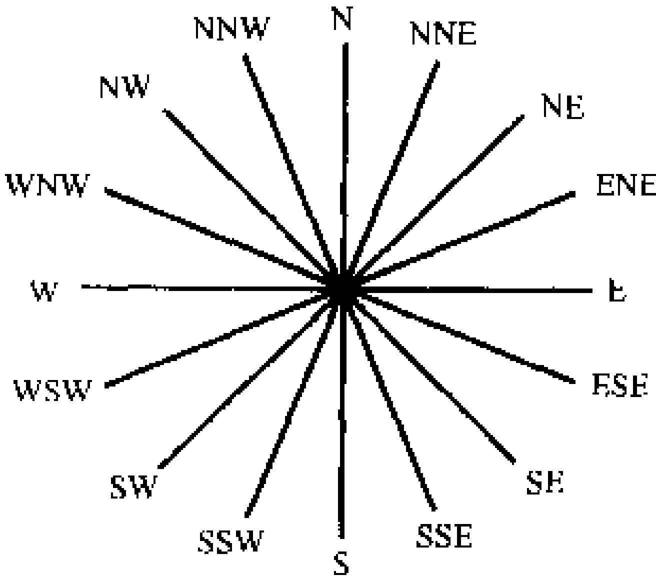


图5-2　16方向描述方法

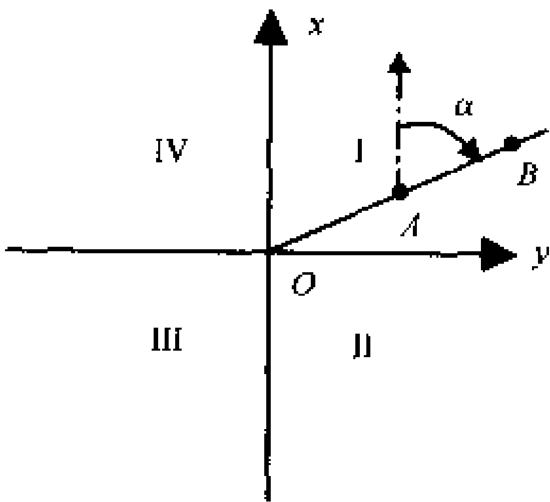


图5-3　空间方向的方位描述

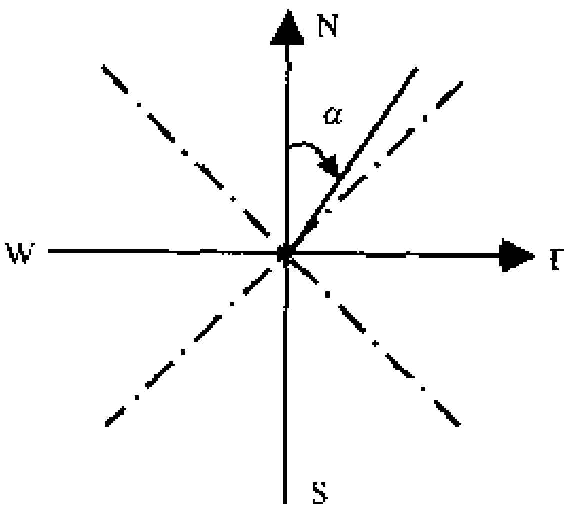


图5-4　空间方向的象限描述

2．空间方向关系描述

所谓空间方向关系，是指在某一参考框架（也称方向参考系统）之中，两个空间目标之间互为源目标和参考目标的相互指向的对等关系。Retz-Schmidt（1988）将参考框架分为3类，如图5-5所示：

（1）内部参考框架（intrinsic reference frame）：即以某一空间目标自身的内部建立的方向参照系统，多用前、后、左、右等方向介词进行描述；

（2）直接参考框架（direct reference frame）：即以观测者所在位置建立的方向参照系统，也多用前、后、左、右等方向介词进行描述；

（3）外部参考框架（extrinsic reference frame）：主要指就地球范围而言，通过选择

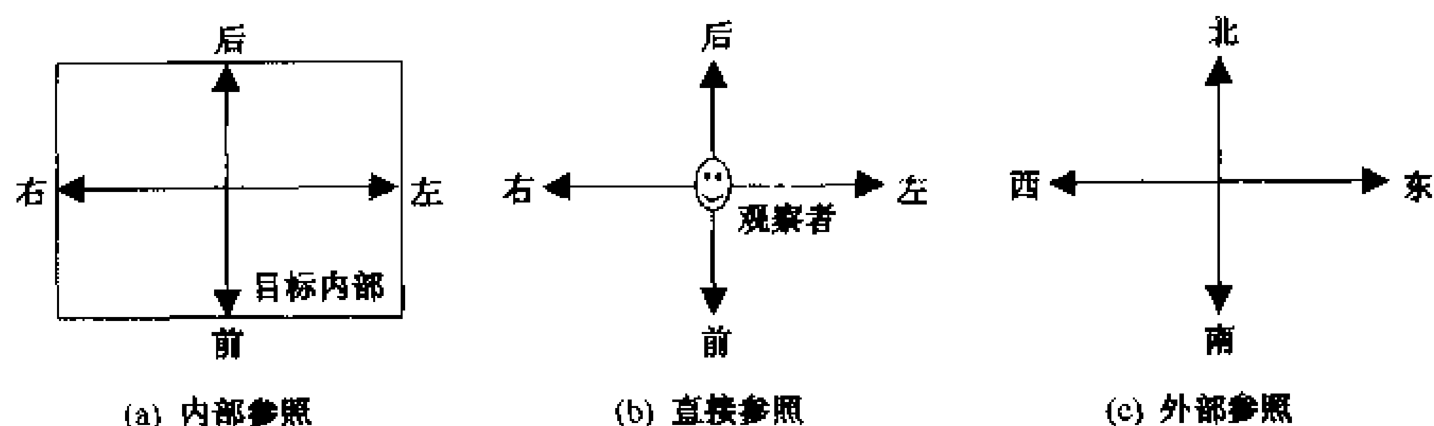


图 5-5　空间方向的方位描述

不同的北方向（如磁北、真北），并经由二维投影转换得到的由东、南、西、北 4 个主方向描述的方向系统。

内部参考框架主要用于小尺度地理空间或建筑、工程领域的方向描述；直接参考框架则为人们日常生活和语言交流所运用；外部参考框架则主要用于学术研究和科技环境。本节的后续讨论将以外部参考框架为主进行。

3．空间方向关系分类

闫浩文（2001）在问卷调查的基础上，从空间认知角度出发将空间方向区分为第一、第二和第三类空间方向关系。

设有两个二维目标（点、线或面））均为非空点的集合，其中之一选为源目标 p，另一选为参考目标 r。并限定：点目标只定位，无大小；线目标连续、可闭合，但不自交；面目标为简单多边形或其嵌套，为闭集，包括面的边界和内部。

(1) 第一类空间方向关系：对应于 $p \cap r = r$ 或 $p \cap r = p$，即两目标是拓扑包含或共位关系。一般认为，在大尺度地理环境中，这种空间方向描述是无意义的，所以又称无意义的空间方向。

(2) 第二类空间方向关系：对应于两目标拓扑相离和拓扑相邻两种情形，是参考目标 r 整体相对于源目标 p 整体的一种方向关系，又称整体对整体的空间方向。

(3) 第三类空间方向关系：对应于两目标（不包括点目标）相交和相互穿越的情形。由于目标在相交和相互穿越过程中，在视觉上造成目标被交点分割为多个子目标，因此在描述两目标的空间关系时，需要结合各子目标的具体情况进行分析。因而，该类空间方向关系称整体对部分的空间方向。

闫浩文（2001）认为，由于上述三类空间方向关系分别对应拓扑包含与共位关系、拓扑相离与相邻关系、拓扑相交关系，而这些拓扑关系是互斥的（即排他的），而且可

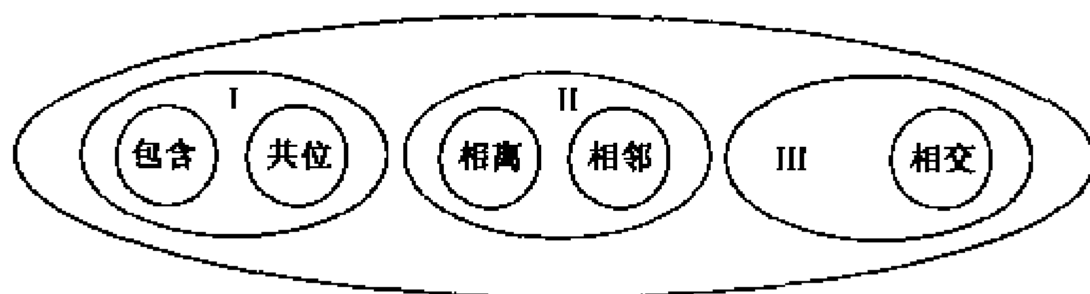


图 5-6　空间方向关系分类的完备性与排他性

（据闫浩文，2001）

以看成拓扑关系的全集（即拓扑关系的其他提法可以归结为此三类），因而是完备的。所以，上述三类空间方向关系具有排他性和完备性，如图 5-6 所示。

5.2.2 空间方向关系描述模型

空间方向关系模型是研究空间方向关系的核心问题，不同的描述模型决定不同的空间关系数据库的组织、检索与推理操作。迄今为止，空间关系描述模型一直是一个相对活跃的研究议题，一方面研究提出了多种空间关系描述模型，另一方面没有形成广泛接受的或商用的描述模型。本节主要介绍 Cone（锥）模型、2D-string（2D-字符）模型、MER（最小外接矩形）模型、MBR（最小约束矩形）模型、FSIA（四半无限区域）模型、DRM（方向关系矩阵）模型和方向 Voronoi 模型，并对其进行简要的比较分析。

1. Cone（锥）模型

Cone 模型最早由 Haar（1976）提出，初衷是探测某一目标是否在某一给定的方向上。其基本思想是：以参考目标 R（Reference，通常为面积较大或较引人注目的目标）的质心 C（Centriod）出发的两条相互垂直的直线将平面区域划分为 4 个无限的锥形（Cone）区域，每个锥的角平分线分别指向东、西、南、北 4 个主方向。然后判断其他目标 T（Target，源目标）分别落在哪一个或哪几个相邻的锥形区域中，即得到源目标与参考目标之间的空间方向关系。如图 5-7 所示。

对于两个拓扑相离的目标，当其距离相对于其尺寸而言较大时，该模型一般能得出正确的结论；反之，或者当目标形状复杂（如相交、缠绕、马蹄形）时，结论可能是错误的，如图 5-8 所示。

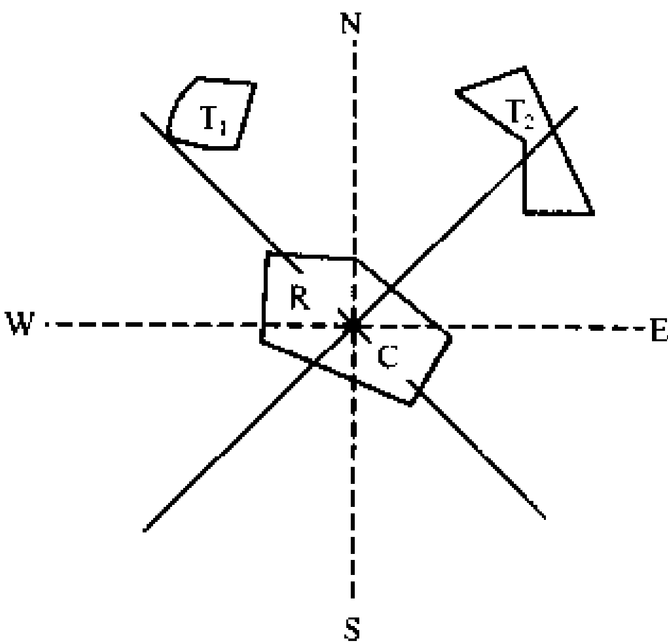


图 5-7 Haar 的 Cone 模型

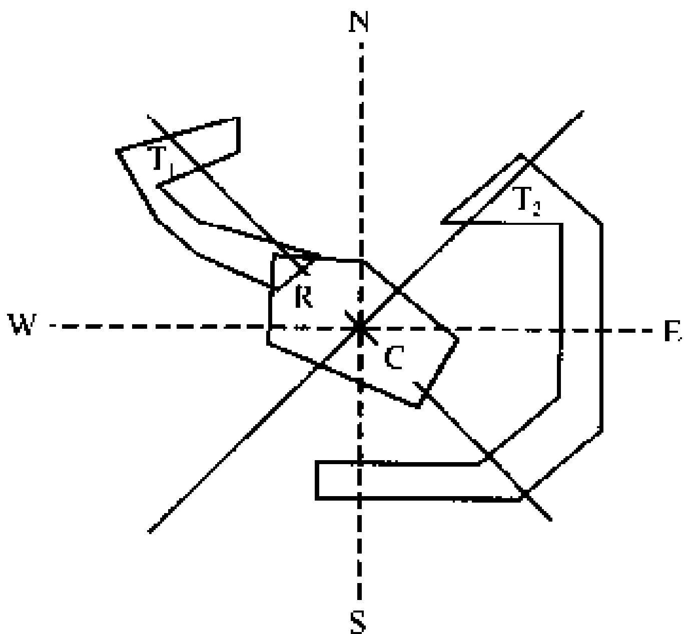


图 5-8 Haar 的 Cone 模型复杂情形

为了克服这一缺陷，Peuquet D. 等（1987）通过引入最小边界矩形（Minimum Boundary Rectangle，MBR）和朝向侧（face side）的概念，对 Cone 模型进行了改进。所谓朝向侧，指参考目标的 MBR 朝向源目标的侧边。通过朝向侧的两个顶点对称地引两条射线，以射线在 MBR 内部的交点取代质心，即锥形可以前后移动（相当于锥形张

角可以变化)。当两个目标相互缠绕时，则用从源目标出发的投影线来判断方向，如图5-9所示。

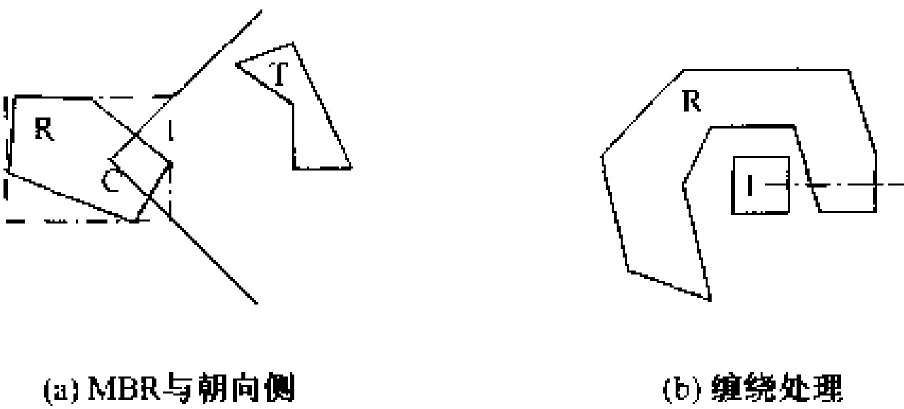
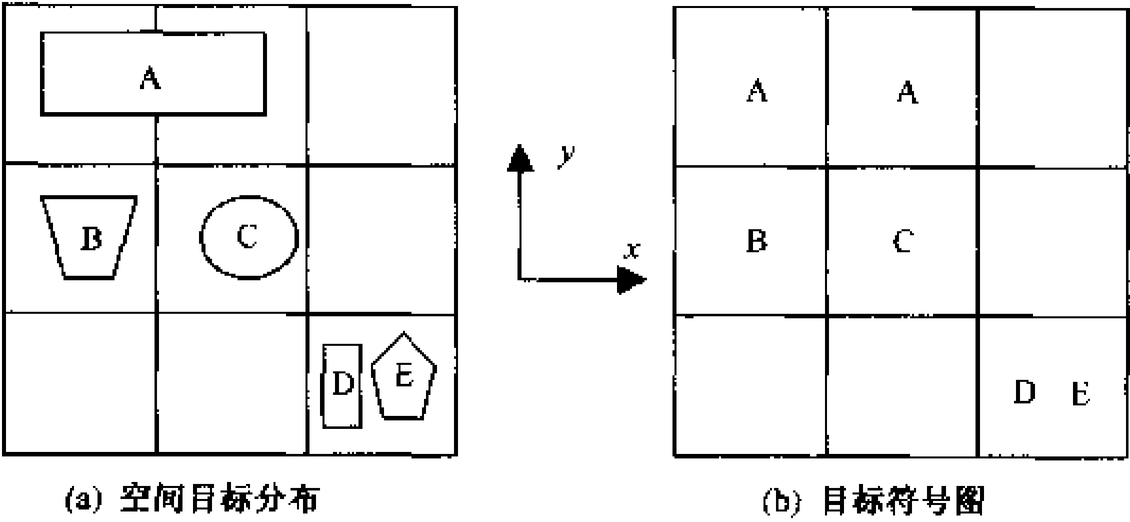


图 5-9　Peuquet D. 的 Cone 模型

2. 2D-string（2D 字符）模型

2D-string 模型最早由 Chang S.K. 等(1987)提出。其实质是一种基于坐标轴的投影,即用固定大小的格网覆盖整个区域,并记录每个格网中的目标标识符号;然后以操作符{=,<,:}("="表示"相等";"<"表示"在前";":"表示"在同一格")顺序记录目标的符号集合Σ。如图5-10所示,Σ在 x、y 轴上的 1D-string 分别为{$A=B<A=C<D:E$}和{$D:E<C=B<A=A$}。将两者合并,即得到该图 2D-string 为:{$A=B<A=C<D:E, D:E<C=B<A=A$}



(a) 空间目标分布　　(b) 目标符号图

图 5-10　2D-string 模型

后来,Chang S. K. 等(1996)通过引入切割线划分机制和 5 元组表示,对上述 2D-string 模型进行了扩展,使最终的空间方向关系表达得更为精确。但是,2D-sring 模型并没有直接计算目标的空间方向关系,而是通过对目标符号图数据库的检索来进行实时的繁琐计算,而且同样无法处理缠绕、包含、相交等复杂情况。

3. MER(最小外接矩形)模型

MER(Minimum Enclosing Rectangle)模型最早由 Mukerjee A. 等(1990)提出。其实质是用空间目标的最小外接封闭矩形(MER)代替空间目标,来对两目标的空间距离关系和空间方向关系同时进行定性描述。其步骤为:①首先获得两空间目标的 MER;②以

MER 的中心代替相应的空间目标;③计算两中心的方向,并相对于参照目标将方向值转化为左前、左后、右前和右后,分别用 1、2、3 和 4 记录下来。

该模型通过对空间目标的抽象近似表示,简单灵活。但对于一些复杂问题的处理,仍然会存在一些问题,如图 5-11 所示,均造成了 T 在 R 左前的错误。

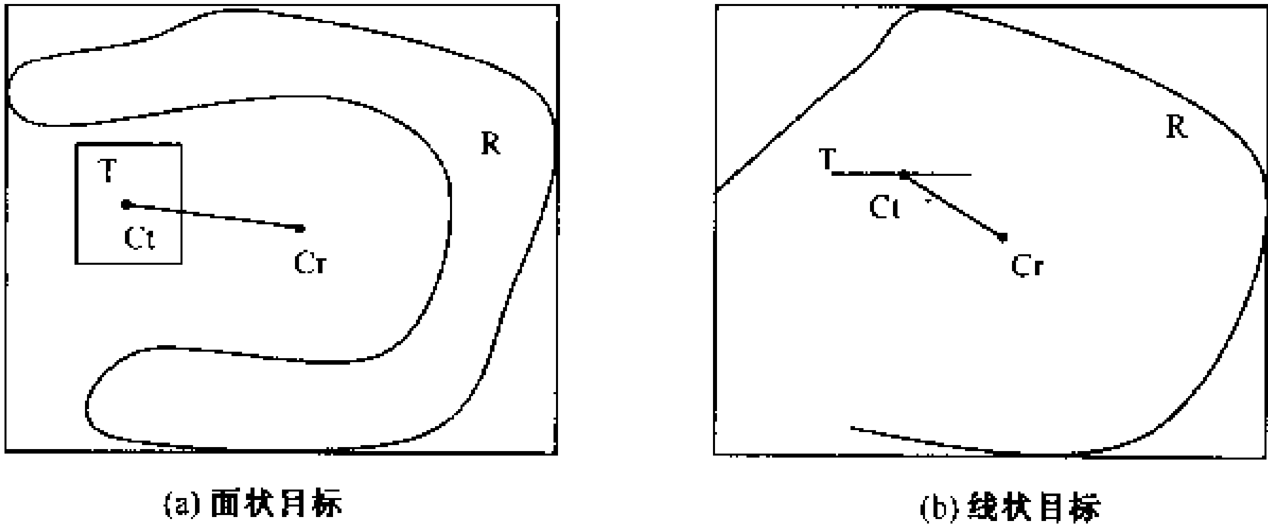


图 5-11　MER 模型处理复杂问题时出错

4. MBR(最小约束矩形)模型

MBR(Minimum Bounding Rectangle)模型由 Papadias D. 等(1994, 1995)明确提出。其实质是通过找出空间目标在 X、Y 轴上投影的最大、最小坐标值,求出由坐标对($X_{\min}$、$Y_{\min}$)、($X_{\max}$、$Y_{\max}$)所构成的约束矩形——即空间目标的 MBR,进而以两空间目标的 MBR 的空间方向关系代替两目标的空间方向关系。可见,MBR 不同于 MER,MBR 基于投影的极值,而 MER 基于图形的最小封闭,因此,前者也称为投影模型。

MBR 模型是基于点集拓扑学的。设平面内两个目标 T、R 分别为源目标和参考目标,MBR-T 和 MBR-R 分别为其 MBR,MBR-R 将平面区域分为如图 5-12 所示的 9 个区域。在这 9 个区域的基础上进行归类合并,得到 North、Same-level、South、West、Same-width 和 East 共 6 种方向关系。

Papadias D. 等(1994, 1995)研究指出,由于两个空间目标在某一坐标轴上的投影最

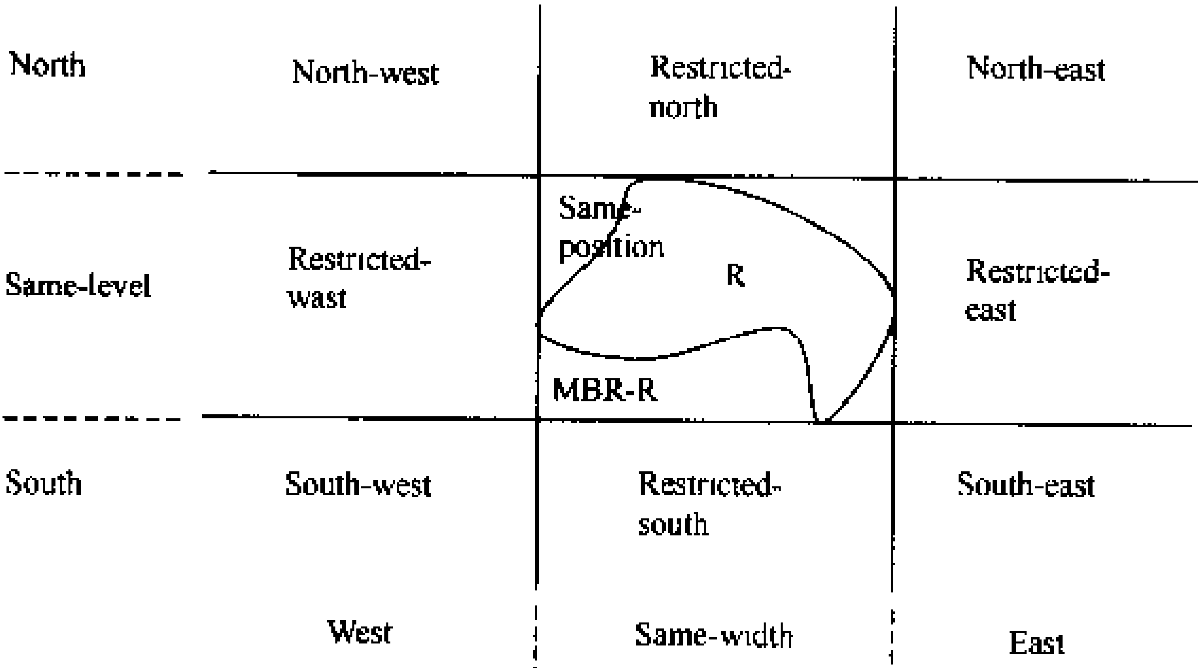


图 5-12　MBR 的方向关系定义

多存在13种可能关系，故MBR-T和MBR-R之间存在13×13=169种可能的位置关系，构成一个13×13的方向关系矩阵。该矩阵包含了Equal、Contain、Inside、Cover、Covered by、Disjoint、Meet和Overlap等8类拓扑关系。Papadias D.等(1994，1995)、Frank A. U.(1996)和Shekhar S.等(1998)基于MBR模型，分别建立了空间方向关系数据存储的B树、B+树和R树模型，并对方向关系的数据库构建、查询以及空间方向推理等进行了研究和探索。

矩形模型简单、直观，当两个空间目标的MBR不相交且描述精度不超过8个主方向时，能得到较满意的结果。但实际上，当两个空间目标邻近时，其MBR相交的频率很高，MBR难以得到满意的结果；同样，MBR模型也基本上无法处理复杂目标的缠绕和包含情况。

5. FSIA(四半无限区域)模型

FSIA(Four Semi-Infinite Area)模型由Abdelmoty A. I.等(1994)提出。该模型本质上与Cone模型一致，主要差别在于方向区域的划分。Cone模型是以参考目标的质心为基点，将平面区域划分为4个锥形的半无限区域；而FSIA模型没有基点，直接以过MBR顶点的4条方向线(NE、NW、SE、SW)及其交点的连线L(统称为5条特征线)将平面区域划分为4个半无限区域(两个尖顶锥形、两个平顶锥形；当MBR为正方形时，均为尖顶锥形)，如图5-13所示。

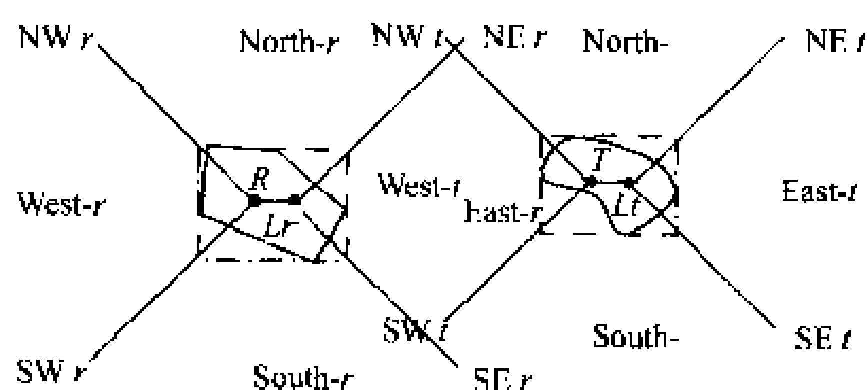


图5-13 FSIA模型

根据FSIA模型的参考目标和源目标各自的5条特征线可以构建如下的空间方向关系矩阵Dir(r, t)：

$$\mathrm{Dir}(r,t)=\begin{bmatrix} NEr\cap NEt & NEr\cap NWt & NEr\cap SEt & NEr\cap SWt & NEr\cap Lt\\ NWr\cap NEt & NWr\cap NWt & NWr\cap SEt & NWr\cap SWt & NWr\cap Lt\\ SEr\cap NEt & SEr\cap NWt & SEr\cap SEt & SEr\cap SWt & SEr\cap Lt\\ SWr\cap NEt & SWr\cap NWt & SWr\cap SEt & SWr\cap SWt & SWr\cap Lt\\ Lr\cap NEt & Lr\cap NWt & Lr\cap SEt & Lr\cap SWt & Lr\cap Lt \end{bmatrix} \tag{5-1}$$

FSIA模型尽管给出了矩阵计算模型，仍不可避免出现Cone模型一样的错误，只不过错误率可以得到一定抑制。Abdelmoty A. I.等(1994)明确指出，该模型不考虑两目标相互缠绕和交叠的复杂情况。

6. DRM(方向矩阵)模型

DRM(Direction Relation Matrix)由 Goyal R. K.(2000)提出。DRM 模型的实质是在 MBR 模型的 9 方向区域划分的基础上，认为源目标 T 总会落在其中某一或某几个相邻的区域中，如图 5-14 所示。从而为参考目标 R 和源目标 T 构建一个 3×3 阶的空间方向关系矩阵 Dir(r,t)：

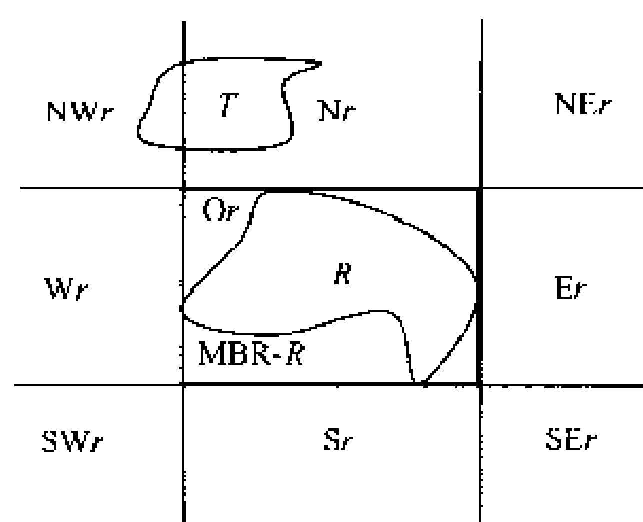


图 5-14 DRM 模型

$$\mathrm{Dir}(r,t)=\begin{bmatrix} \mathrm{NW}r\cap T & \mathrm{N}r\cap T & \mathrm{NE}r\cap T \\ \mathrm{W}r\cap T & \mathrm{O}r\cap T & \mathrm{E}r\cap T \\ \mathrm{SW}r\cap T & \mathrm{S}r\cap T & \mathrm{SE}r\cap T \end{bmatrix} \tag{5-2}$$

当源目标 T 与 9 方向区域之一有重叠时，则交集非空(取为 1)；否则为空(取为 0)。Goyal R. K. 分析认为，矩阵 Dir(r,t)有 $2^9=512$ 种取值可能，其中 218 种有意义；进而为该矩阵建立了一种类似于图像的模板集合。为了提高 DRM 模型的方向描述精度，Goyal R. K. 引入面积比例系数(即源目标落在某一方向区域的面积与其总面积之比)来代替原矩阵中的非空值(1)：

$$\mathrm{Dir}(r,t)=\begin{bmatrix} \dfrac{\mathrm{Area}(\mathrm{NW}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{N}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{NE}r\cap T)}{\mathrm{Area}\,T} \\ \dfrac{\mathrm{Area}(\mathrm{W}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{O}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{E}r\cap T)}{\mathrm{Area}\,T} \\ \dfrac{\mathrm{Area}(\mathrm{SW}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{S}r\cap T)}{\mathrm{Area}\,T} & \dfrac{\mathrm{Area}(\mathrm{SE}r\cap T)}{\mathrm{Area}\,T} \end{bmatrix} \tag{5-3}$$

Goyal R. K. 还讨论了当参考目标为点、垂直线段、水平线段等特殊情况下方向关系矩阵的建立；并对源目标位于方向划分线之上的特殊情况进行了相应讨论。曹菡等(2001)进一步对基于该模型的空间方向定性描述和推理方法进行了扩展研究，提出了一种空间方向关系描述的分层多级处理方法，建立了以点、线、面为参照目标的方向关系描述的 3 层模式结构，其中每层模式由基本方向关系描述、推理模型和方向关系描述、推理细化模型组成。

7. 方向 Voronoi 模型

闫浩文(2001)基于心理学的实验论证，认为"决定两目标之间的空间方向关系的主要因素是这两个目标之间的可视区域，而可视区域恰好是它们的邻近区域，因而两目标的空间方向计算问题可以归结为目标间邻近区域内的问题"。考虑到 Delaunay TIN 和 Voronoi 图是处理空间邻近问题的优良工具，闫浩文利用 Voronoi 图、Delaunay TIN 和凸壳等计算几何的概念和算法，研究提出了空间方向关系描述的 Voronoi 模型。由于大部分空间目标的 Voronoi 图是多条线段连成的折线，因此空间目标之间的方向关系就需要多次描述，多次描述的并集即是对空间目标方向关系的完整描述，这就是方向 Voronoi 模型中的"先分治、后合治"思想。

从空间目标的图形信息到定性的空间方向描述是一个信息过滤的过程，两个空间目标的方向关系的判断则是人的空间认知过程。闫浩文借用认知心理学的 Triesman-Broadbent 过滤器设计了这一空间认知过程的模型框架，如图 5-15 所示。

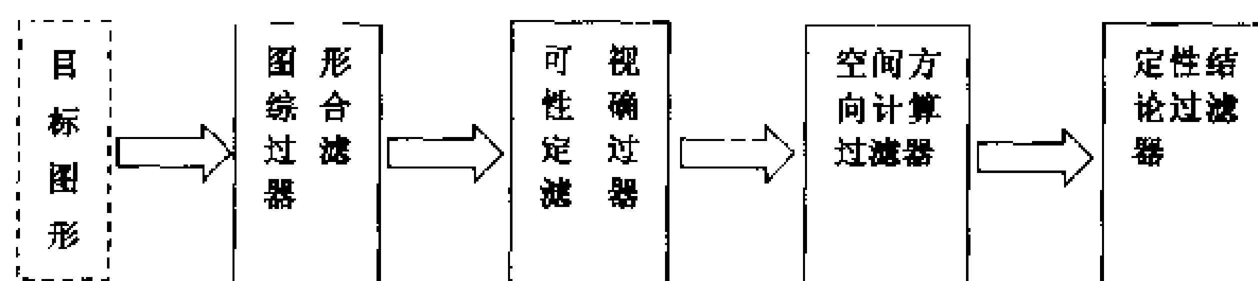


图 5-15　空间方向判断过滤器模型框架
（闫浩文，2001）

（1）图形综合过滤器：首先对两个目标的拓扑关系进行判断和预处理，然后从中提取图形特征信息（群点凸壳直径、特征弯曲等），确定图形综合幅度，并进行图形综合。

（2）可视性确定过滤器：从综合后的目标图形中提取视点，形成可视链，构造可视区域的 Delaunay TIN。

（3）空间方向计算过滤器：接受可视区域 Delaunay TIN 信息，生成两目标的方向 Voronoi 图；然后结合目标和 Voronoi 边的侧位关系，计算 Voronoi 边的方位角。

（4）定性结论过滤器：按人们习惯的方式表达计算结果。有两种方式，其一按可视区域内 Voronoi 边在 8 方向域中的长度分别占 Voronoi 边总长度的比例进行详细表示；其二按可视区域内首尾 Voronoi 点的连线的方位角进行概略表示。如图 5-16 所示。显然，前者可以描述任意复杂情况，而后者则对于包围、缠绕等复杂情况结论不一定正确。

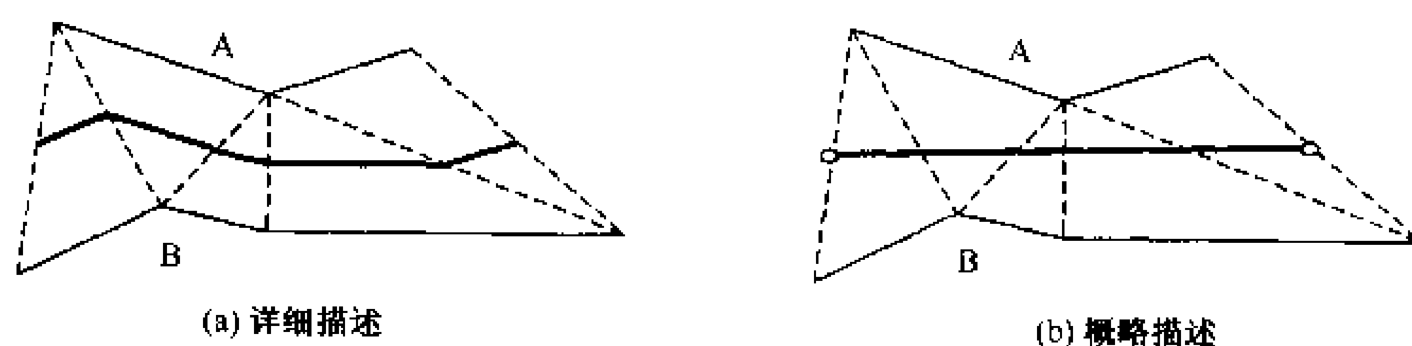


图 5-16　基于 Voronoi 模型的空间方向描述
（闫浩文，2001）

由于两空间目标的方向 Voronoi 图总是存在的且是惟一的，因此方向 Voronoi 模型描述的空间方向关系具有完备性和惟一性；而且无需进行参考目标与源目标的区分，具有普遍性。该模型与 Cone、MER、MBR、2D-string 等模型相比，模型的复杂性有所增加，计算过程比较繁琐。

8．模型的形式化比较

从形式化要求的完备性、惟一性、普遍性和严密性角度考虑，可对上述 7 种模型进行如表 5-1 所示的比较。需要再次指出，目前尚没有哪一种描述模型得到普遍接受，在被商业 GIS 软件采用之前，都还有一段较长的路要走。

表 5-1 空间方向描述模型的形式化程度对比

描述模型	完备性	惟一性	普遍性	严密性
CONE 模型	不研究相邻、相交等关系	能保证	差	复杂情况出错
2D-string 模型	好	能保证	好	复杂情况出错
MER 模型	一般	能保证	差	复杂情况出错
MBR 模型	较好	能保证	好	复杂情况出错
FSIA 模型	不研究相邻、相交等关系	能保证	差	不考虑复杂情况
DRM 模型	好	能保证	一般	基本正确
方向 Voronoi 模型	好	能保证	好	较好

§5.3 空间距离关系

空间距离是最基本、最重要的一类空间关系，空间方向关系、空间拓扑关系和空间相似关系、相关关系的研究都离不开空间距离这一基础，缓冲区分析、网络分析如最短路径分析、推销员问题等都是以空间距离为基础的。

空间距离关系包括点-点、点-线、点-面、点-体、线-线、线-面、线-体、面-面、面-体和体-体共 10 类。根据 GIS 所采用的数据结构不同，空间距离的计算有基于二维和三维欧氏空间的矢量距离和基于栅格的栅格距离。对于点-点空间距离关系而言，其空间距离计算比较简单；而对于其他 9 类空间距离关系的空间距离计算则较为复杂，尤其是面-面、面-体和体-体空间距离关系，其空间距离的描述和计算都相当复杂。为了计算和分析空间不相交实体之间的距离关系，人们定义了最近距离、最远距离和平均距离等概念，如图 5-17所示。

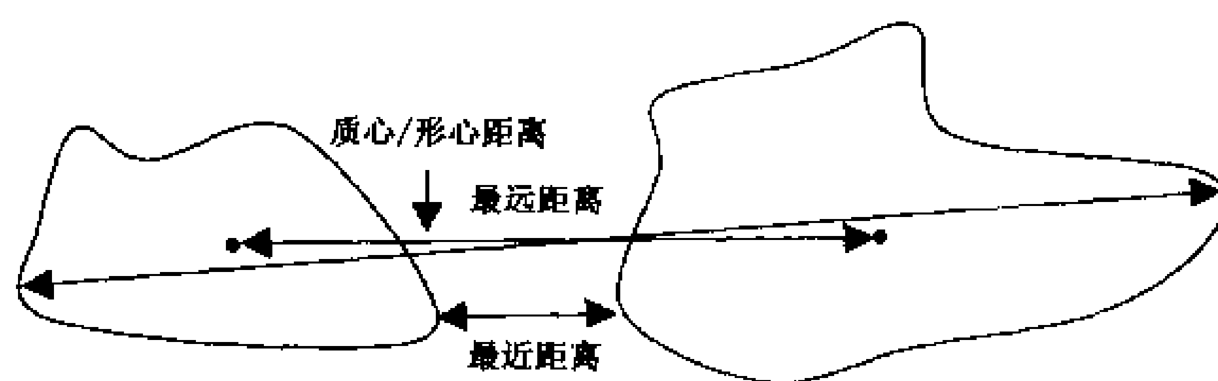


图 5-17 空间距离的基本定义

所谓最近距离，一般定义为两空间目标的最邻近点之间的距离。对于点-直线距离这一特例，如图 5-18 所示，其最近距离为点到直线的垂直距离或到直线较近端点的距离；对于直线-直线空间距离关系这一特例而言，如图 5-19 所示的两条不相交直线，其最近距离 d_{min}可用下式描述：

$$d_{min} = \min(ac, ad, af, bc, bd, de) \qquad (5\text{-}4)$$

最近距离概念在森林防火中有重要意义，即任何火源(点)距离森林(面)必须大于某一安全临界值。

所谓最远距离，定义为两空间目标的最远离点之间的距离。在电视转播和移动通信

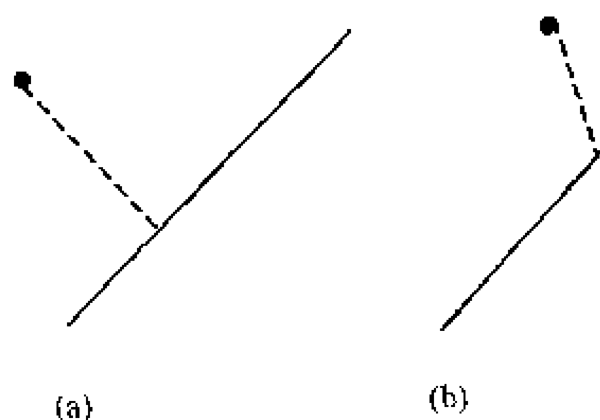


图 5-18 点-直线的最近距离

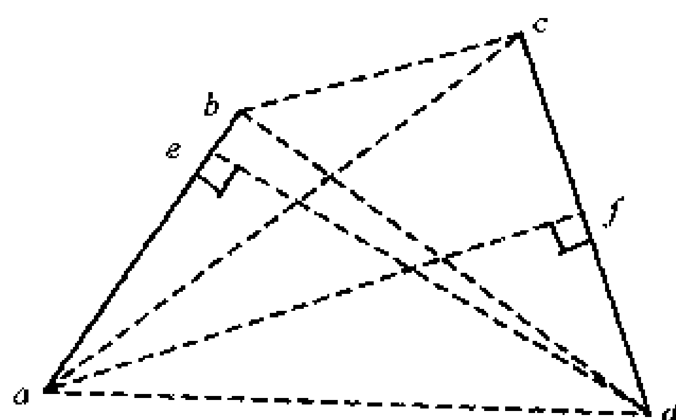


图 5-19 直线-直线的最近距离求解

覆盖范围分析中，为了保证区域内的任何点位均能接收到信号，必须使用最远距离。

所谓平均距离，则其定义和理解多种多样。有的将其定义为两空间目标的所有点对之间的距离的平均值（要求所有点配对，不完全配对时需进行内插）；有的将其定义为最近距离与最远距离的平均值；有的则将其定义为两空间目标的质心或形心之间的距离。作者认为，上述定义存在一个共同问题，即将两目标的整体纳入分析，而没有考虑目标之间的可视性问题。事实上，人们在认识客观世界时，关心的是视线所及部分或客观可达部分。比如：当一个人站在某一座大山之前目测他与大山的距离时，并不考虑大山到底有多大，以及其背后延伸多远。因此，作者建议：从人的认知习惯出发，应以两目标之间的可视区域为基准，分别研究其最近距离、最远距离和平均距离。

§5.4 2D空间拓扑关系

"拓扑"一词来自于希腊文，意思是"关于形状的研究"。拓扑关系（Topological relationship）反映了空间目标的逻辑结构，对空间目标查询、分析和空间目标重建具有重要意义。拓扑关系是指拓扑变换下的拓扑不变量（如多边形内部的点），即不考虑度量和方向的空间目标之间的关系，包括相邻（adjacent）、邻接（connection）、关联（conjunction）和包含（inclusion）等。

5.4.1 点集拓扑学原理

拓扑学是研究图形在拓扑变换下的不变性质，已成为GIS空间关系的理论基础，为空间关系的描述提供直接的理论依据。以点集拓扑学为例，其主要定义定理如下：

定义1：X 为一非空集合，$\rho: X\times X\rightarrow R$ 为一映射，如果对于任意的 $x, y, z\in X$ 有：

(1) $\rho(x,y)\geqslant 0$，并且 $\rho(x,y)=0$ 当且仅当 $x=y$；

(2) $\rho(x,y)=\rho(y,x)$；

(3) $\rho(x,z)\leqslant\rho(x,y)+\rho(y,z)$（三角不等式）；

则称 ρ 为 X 的度量，偶对 (X,ρ) 称为度量空间。

定义2：A 为度量空间 X 的子集，如果 A 的每一个点都有一个球形邻域包含于 A，则称 A 为 ρ 的开集。

定义3：X 为非空集合，A 为 X 的子集族，如果满足下列条件：

(1) X 和空集$\varnothing$属于 A；

(2) 若 A_1、A_2 属于 A，则 A_1、A_2 的交集属于 A；

(3) A 中任意两个元素的并仍为 A 中的元素；

则称 A 为 X 的拓扑，偶对(X, A)为拓扑空间。

定义 4：A 为拓扑空间 X 的子集，如果点 x 属于 X 的每一邻域 U 中都有 A 中异于 x 的点，即 $U \cap (A - \{x\}) \neq \varnothing$，则称 x 为 A 的聚点或极限点。A 的聚点可以属于 A 也可以不属于 A。

定义 5：A 为拓扑空间 X 的子集，集合 A 的所有聚点构成的集合称为 A 的导集，记作 $dX(A)$ 或 $d(A)$。

定义 6：A 为拓扑空间 X 的子集，如果 A 的每一聚点都属于 A，即 $d(A)$ 为 A 的子集，称 A 为闭集。

定义 7：X 为拓扑空间，X 的子集 A 与 A 的导集 $d(A)$ 的并集 $A \cup d(A)$ 称为 A 的闭包，记作 $C(A)$。

定义 8：A 为拓扑空间 X 的子集，如果 A 是点 x 属于 X 的邻域，即存在 X 的开集 U 使得 x 属于 U，U 为 A 的子集，则称点 x 为集合 A 的内点。集合 A 的所有内点构成的集合，称为 A 的内部，记作 $I(A)$。

定义 9：A 为拓扑空间 X 的子集，对于点 x 属于 X，如果存在 x 的任一邻域 U 中既有 A 的点又有 $\sim A$ 的点，即 $U \cap A \neq \varnothing$ 且 $U \cap (\sim A) \neq \varnothing$，则称 x 为集合 A 的边界点。集合 A 的所有边界点的集合称为集合 A 的边界，记作 $B(A)$。

定理 1：A 为拓扑空间 X 的任一子集，则：

(1) $C(A) = \sim I(\sim A) = I(A) \cup B(A)$；

(2) $I(A) = \sim C(\sim A) = C(A) \cup \sim B(A)$；

(3) $B(A) = C(A) \cap C(\sim A) = \sim [I(A) \cup I(\sim A)] = B(\sim A)$。

5.4.2 2D空间拓扑关系描述

以二维 GIS 为例，描述零维、一维、二维空间目标之间的空间拓扑关系通常采用拓扑相邻、拓扑邻接、拓扑关联和拓扑包含 4 类空间关系。若将具有相同维数的空间目标称为同维目标，则同维目标中，按空间目标的重要性可以进一步分为不同等级，如包含其他面域的面域的等级比被包含的面域的等级高。据此，拓扑相邻、拓扑邻接、拓扑包含和拓扑关联可定义如下：

(1) 拓扑相邻，是指同维或不同维目标之间的相互邻近（可视）而不相互接触（靠近）的关系，如零维的电线杆或树木与 1 维的道路邻接；1 维的高速公路与 1 维的铁路相邻；2 维的池塘与 1 维的道路相邻；2 维草地与 2 维建筑相邻等，如图 5-20 所示。Christopher G. M.（1995）基于空间目标 Voronoi 图的定义和计算，将空间邻近（spatial adjacency）概念定义为具有公共 Voronoi 边的两个空间目标之间的关系，并指出空间存在 k 阶邻近关系。胡勇等（1997）进一步定义了最邻近、次邻近概念。

(2) 拓扑邻接，是指同维目标之间的相互邻接关系，如零维结点通过 1 维弧段相互邻接；1 维弧段通过 0 维结点相互邻接；2 维面域通过 1 维弧段相互邻接。

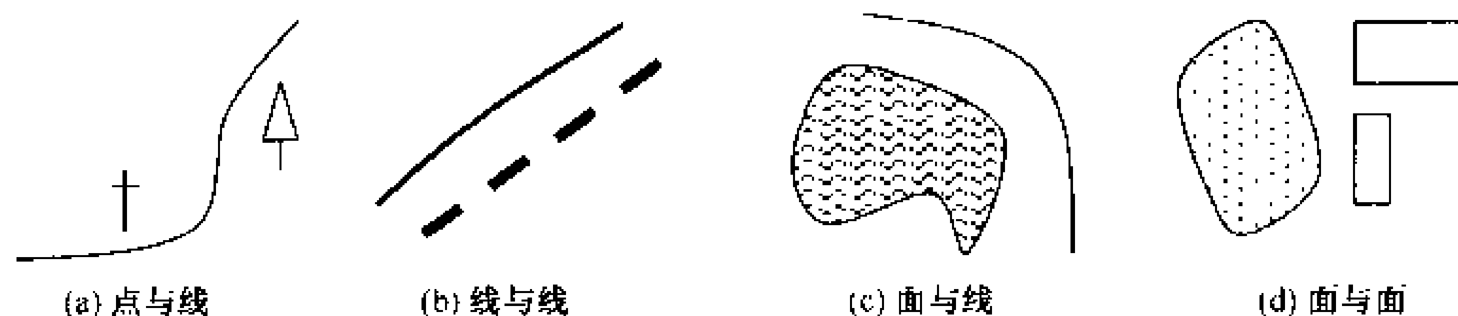


图 5-20　拓扑相邻关系

（3）拓扑关联，是指不同维目标之间的相互关联关系，如零维结点与 1 维弧段相关联；1 维弧段与 2 维面域相关联。

（4）拓扑包含，一般指同维不同级目标之间的包含关系，主要指高一级别的二维面域目标对低一级别的面域目标的空间包含关系。若将拓扑包含关系扩展，也可以是指高维目标对低维目标的包含关系，如二维面域包含 1 维弧段或 0 维点目标。

通常，拓扑邻接、关联和包含关系在计算机中以拓扑结构表的形式进行描述。如图 5-21 所示的简单地图为例，其拓扑邻接、关联和包含关系可用表 5-2、表 5-3 和表 5-4 表示：

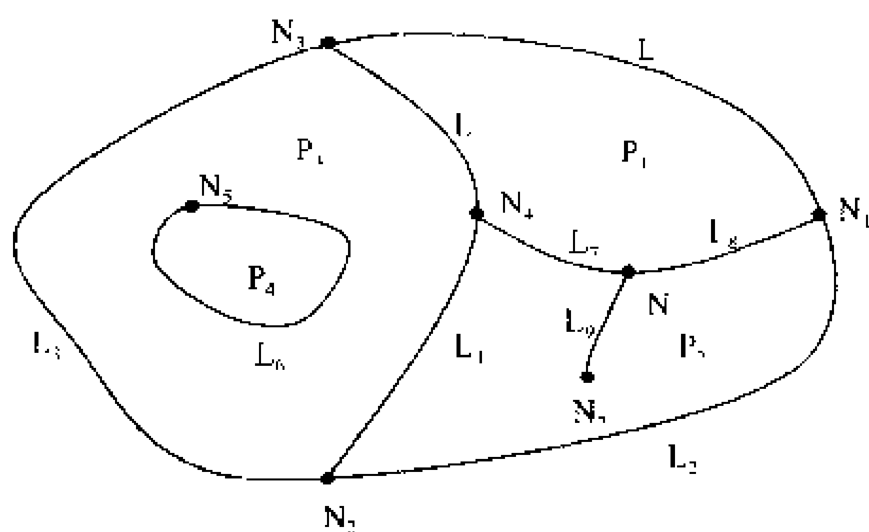


图 5-21　拓扑邻接、关联和包含关系实例

表 5-2a　结点拓扑邻接关系

结 点	邻接结点
N_1	N_2，N_3，N_6
N_2	N_1，N_3，N_4
N_3	N_1，N_2，N_4
N_4	N_2，N_3，N_6
N_5	—
N_6	N_1，N_4，N_7
N_7	N_6

表 5-2b　弧段拓扑邻接关系

弧 段	邻接弧段
L_1	L_2，L_3，L_5，L_8
L_2	L_1，L_3，L_4，L_8
L_3	L_1，L_2，L_4，L_5
L_4	L_2，L_3，L_5，L_7
L_5	L_1，L_3，L_4，L_7
L_6	
L_7	L_4，L_5，L_8，L_9
L_8	L_1，L_2，L_7，L_9
L_9	L_7，L_8

表 5-2c　面域拓扑邻接关系

面 域	邻接面域
P_1	P_2，P_3
P_2	P_1，P_3
P_3	P_1，P_2，P_4
P_4	P_3

表 5-3a 结点-弧段拓扑关联

结点	关联弧段
N_1	L_1、L_2、L_8
N_2	L_2、L_3、L_4
N_3	L_1、L_3、L_5
N_4	L_4、L_5、L_7
N_5	L_6
N_6	L_7、L_8、L_9
N_7	L_9

表 5-3b 弧段-面域拓扑关联

弧段	关联面域
L_1	P_1
L_2	P_2
L_3	P_3
L_4	P_2、P_3
L_5	P_1、P_3
L_6	P_3、P_4
L_7	P_1、P_2
L_8	P_1、P_2
L_9	—

表 5-4a 面域-面域拓扑包含

面域	被包含面域
P_1	—
P_2	—
P_3	P_4
P_4	—

表 5-4b 面域-弧段拓扑包含

面域	被包含弧段
P_1	—
P_2	L_9
P_3	—
P_4	—

5.4.3 2D空间拓扑关系维护

动态地维护GIS中空间对象的拓扑关系是GIS系统维护中一件至关重要而又十分繁杂和费时的工作。以早期的ARC/INFO为例，一幅图中哪怕仅对局部做了很小的修改（比如增加一条线，删除一个面域等）都需要对整幅图重建拓扑，并改写整个数据文件。

如果适当地考虑了结点和弧段的共享问题，并且按结点-弧段的拓扑关联关系设计了相应的拓扑关联表（如3.1.2节所述），则可以为拓扑数据维护带来很多方便。

以图5-22所示的弧段拓扑维护为例：如果事先建立了如表5-3a所示的结点-弧段拓扑关联表和如表3-2所示的结点点位坐标数据文件，当移动或修改某一结点的位置时，仅修改坐标数据文件中相应的点位坐标，其他拓扑关系完全不变。否则，如图5-23所示，所有与此有关的弧段都要逐条修改，所移动的结点与三条弧段相关联，故需要修改

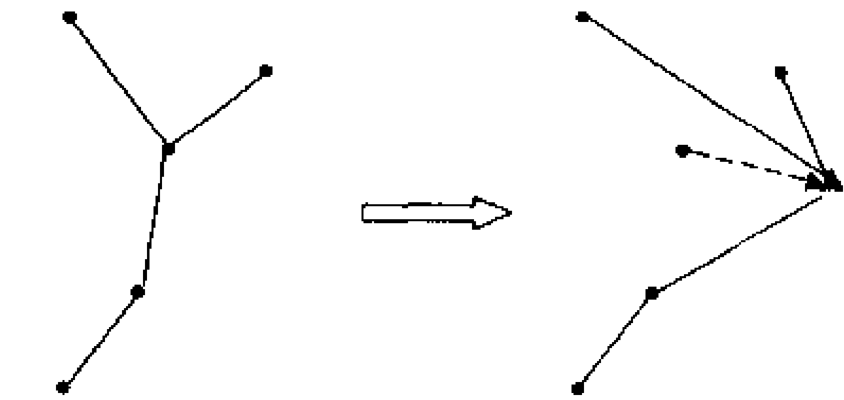


图5-22 建立了弧段-结点拓扑关联关系的结点修改

3 次。显然，与被修改点关联的弧段越多，修改工作越麻烦。

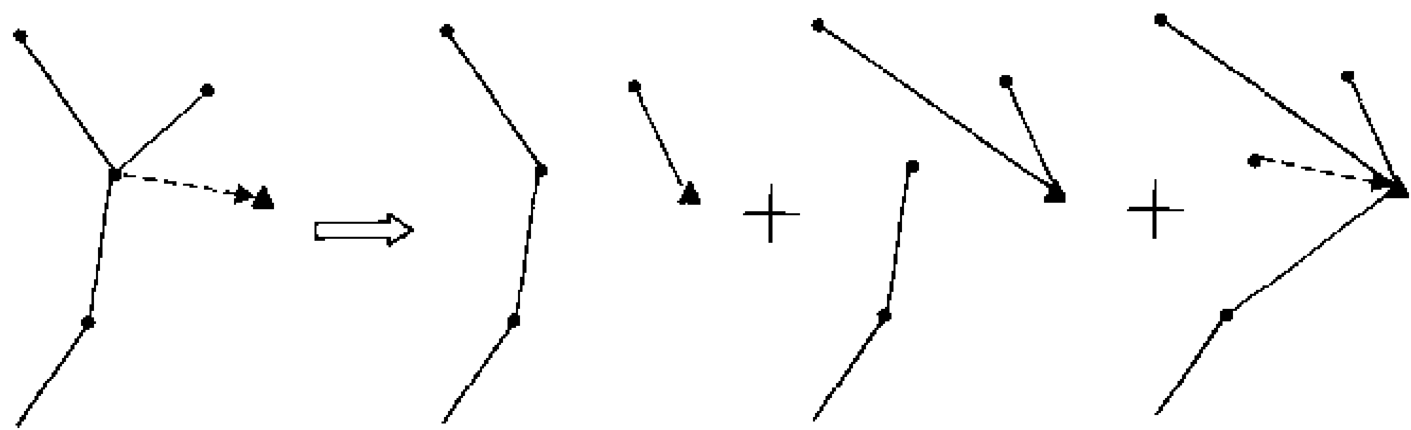


图 5-23　未建立弧段-结点拓扑关联关系的结点修改过程

面域拓扑维护也是同样道理。如果事先建立了如表 5-3b 示的弧段-面域的拓扑关系，则面域公共边的修改和拓扑维护也简单得多。如图 5-24 所示，面域 P_1、P_2 公共边的 2 号结点作了一次修改，只需修改如表 3-2 所示的点位坐标数据文件即可。否则，需要对两个面域的数据分别进行修改，如图 5-25 所示。

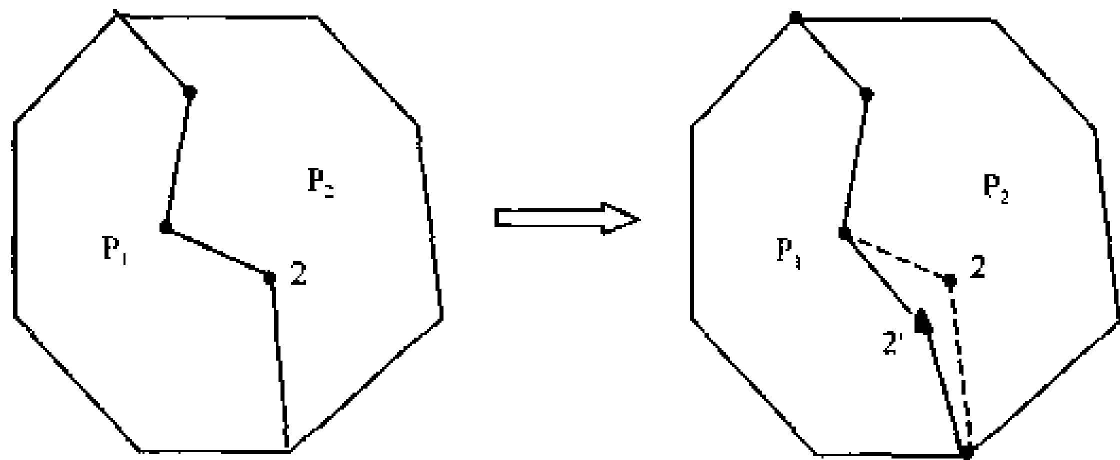


图 5-24　建立了面域-弧段拓扑关联关系的面域修改

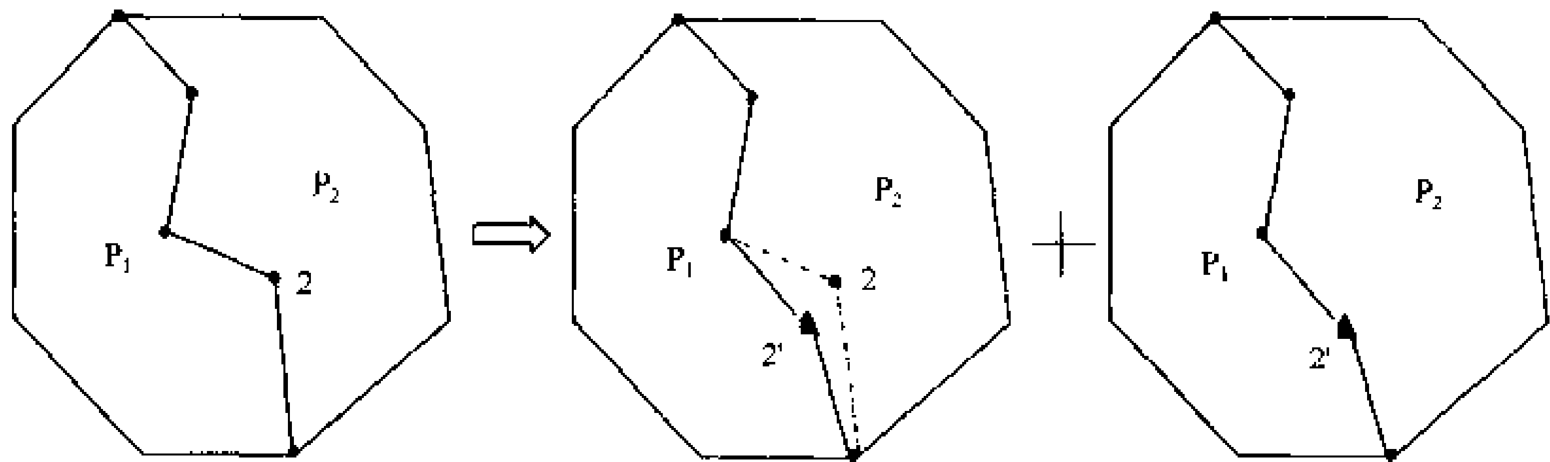


图 5-25　未建立面域-弧段拓扑关联关系的面域修改过程

5.4.4　9I 模型

Egemhofer M. J. 等（1991）基于点集拓扑学提出 4-交（four intersection，4I）模型（也称 4-元组模型）来描述空间拓扑关系。4I 模型基本原理为：设有空间目标 A、B，分别用∂A、A°和∂B、B°表示其边界和内部，则其空间拓扑关系 Topo（A，B）可

用∂A、A°和∂B、B°两个点集的交来定义，即可以用如下一个2×2的矩阵进行描述。

$$\mathrm{Topo}(A,B)=\begin{bmatrix}\partial A\cap\partial B & \partial A\cap B^\circ\\ A^\circ\cap\partial B & A^\circ\cap B^\circ\end{bmatrix}\tag{5-5}$$

矩阵中每个“元组”有“空”（∅）和“非空”（¬∅）两种取值，分别表示相离或相交。因此，4-交模型的Topo（A，B）共有$2^4=16$种可能的取值。排除现实世界中没有物理意义的关系，可以导出8种面-面关系、13种面-线关系、3种面-点关系、16种线-线关系和3种点-线关系。尽管如此，许多人眼很容易识别的空间关系，利用该模型反而难以识别。

针对4I模型的不足，Egemhofer M. J. 等（1995）进一步考虑了空间目标的外部，提出用9-交（nine intersection，9I）模型（也称9-元组模型）来描述空间拓扑关系。设有空间目标A、B，分别用∂A、A°、A^{-1}和∂B、B°、B^{-1}表示其边界、内部和外部，则其空间拓扑关系Topo（A，B）可以用以下矩阵描述：

$$\mathrm{Topo}(A,B)=\begin{bmatrix}\partial A\cap\partial B & \partial A\cap\partial B^\circ & \partial A\cap B^{-1}\\ A^\circ\cap\partial B & A^\circ\cap B^\circ & A^\circ\cap B^{-1}\\ A^{-1}\cap\partial B & A^{-1}\cap B^\circ & A^{-1}\cap B^{-1}\end{bmatrix}\tag{5-6}$$

矩阵中每个“元组”有“空”（∅）和“非空”（¬∅）两种取值，分别表示相离或相交。因此，Topo（A，B）共有$2^9=512$种可能的取值，包括了所有的拓扑关系，具备了理论上的完备性。但实际上，其中的某些关系并不存在或是无用的，有些关系在空间上还可能相互矛盾。因此，在运用9I模型时，需要对拓扑关系进行必要的“提炼”或“过滤”。

表5-5列出了面-面（A-A）、面-线（A-L）、面-点（A-P）、线-线（L-L）、线-点（L-P）、点-点（P-P）之间可能空间关系的矩阵表达，其中：① 9位二进制数表示矩阵的9个取值，1表示交集非空，0表示交集为空；② “—”表示该关系不可能存在；“Yb”表示在单值（地物不重叠）和多值（地物可重叠）的矢量图上都可能存在的关系，“Ym”表示仅在多值矢量图上存在的关系。从表5-5可以看出，可能的拓扑关系数目只有60种（面-面10种、面-线19种、面-点3种、线-线23种、线-点3种、点-点2种），远远少于512。

图5-26所示为上述可能拓扑关系举例。从某种意义上说，9I模型所描述的拓扑关系只是拓扑关系的类别，每一类别中有多种可能的情形：如两条曲线相交，一个交点和多个交点的9I模型表示是一样的，但其拓扑关系并不相同。

表5-5　9I模型表示的两个2D要素可能的拓扑关系

关　系	9I矩阵	A-A	A-L	A-P	L-L	L-P	P-P
R026	000011010	—	—	—	—	—	Yb
R030	000011110	—	—	Yb	—	Yb	—
R031	000011111	Yb	Yb	—	Yb	—	—
R063	000111111	—	Yb	—	Yb	—	—
R092	001011100	—	—	Yb	—	Yb	—
R093	001011101	—	—	—	Yb	—	—
R095	001011111	—	—	—	Yb	—	—

续表

关系	9I 矩阵	A-A	A-L	A-P	L-L	L-P	P-P
R127	001111111	—	—	—	Yb		—
R159	010011111	—	—	—	Yb	—	
R179	010110011	Ym	—	—	Ym	—	—
R191	010111111	—	Yb	—	Yb	—	—
R220	011011100	Ym	Yb	—	Ym	—	—
R223	011011111	—	—	—	Yb	—	—
R252	011111100	—	Yb		—	—	—
R253	011111101	—	Yb	—	—	—	—
R255	011111111	—	Yb	—	Ym		—
R272	100010000		—	—	—	—	Ym
R277	100010101	—	—	—	Yb	—	
R279	100010111	Yb	—	—	—	—	
R284	100011100	—	—	Yb	—	Yb	—
R285	100011101	Yb	Yb	—	—		
R287	100011111	Yb	Yb	—	Yb	—	—
R311	100110111	—	—	—	Yb	—	—
R316	10011110	—	Yb	—	—	—	—
R317	100111101	—	Yb	—	—	—	—
R319	100111111	—	Yb	—		—	—
R349	101011101	—	—	—	Yb	—	—
R373	101110101	—	—	—	Yb	—	—
R400	110010000	Ym	—	—	Ym		
R412	110011100	—	Yb	—	—	—	—
R415	110011111	—	—	—	Yb	—	—
R435	110110011	Ym	—	—	Ym	—	—
R439	110110111	—	—	—	Yb	—	—

5.4.5 2D空间拓扑关系研究进展

赵仁亮等（2000）进一步利用基于 Voronoi 图的 9I 模型（V9I 模型），研究了空间关系的映射与操作，建立了底层数据结构与空间关系语义层之间的 V9I 映射机制，并在 Voronoi 动态栅格算法的基础上，构建了空间关系的基本操作。

王家耀（2001）提出运用维数扩展（dimension extension，DE）法将 9-交模型进行扩展，提出基于维数扩展的 9-交（DE-9I）模型。该模型首先定义几何实体的边界是比其低 1 维的几何实体的集合，即点的边界为空集；非闭合线的边界为其两个端点；闭合线的边界为空集；面的边界为其边界线段的集合。设 P 为某一点集，DIM 为求维函数，则：

$$\mathrm{DIM}(P)=\begin{cases}-1, P \text{ 等于空} \\ 0, P \text{ 不包含线、面,但至少包含 1 个点} \\ 1, P \text{ 不包含面,但至少包含 1 条线} \\ 2, P \text{ 至少包含 1 个面}\end{cases} \tag{5-7}$$

据此，可将式（5-4）改写为 DE-9I 模型的空间关系矩阵：

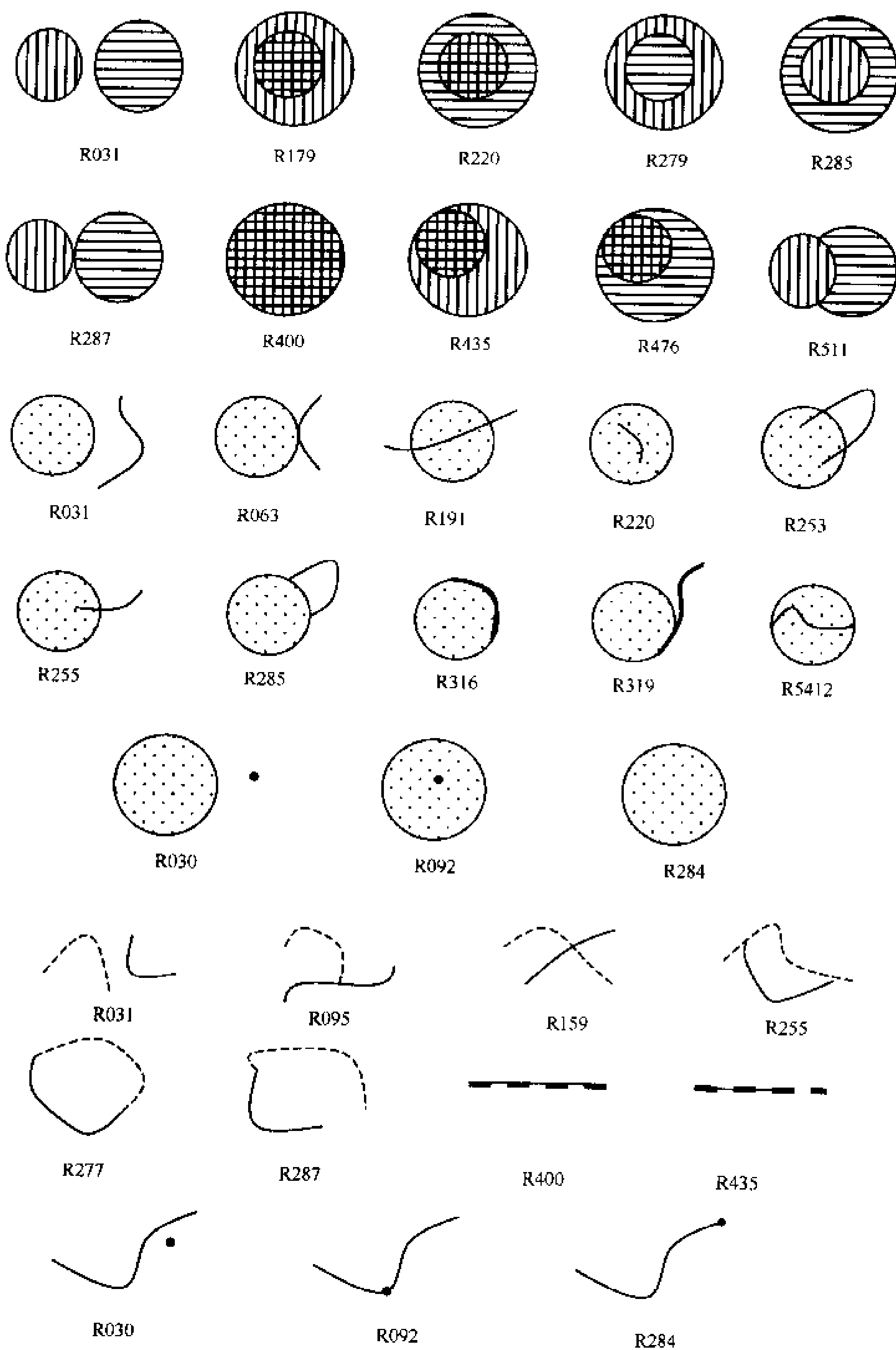


图 5-26 91 模型拓扑关系举例（面-面、面-线、面-点、线-线、线-点）

$$
\mathrm{DIM}(A, B) = \begin{bmatrix} \mathrm{DIM}(\partial A \cap \partial B) & \mathrm{DIM}(\partial A \cap B^{\circ}) & \mathrm{DIM}(\partial A \cap B^{-1}) \\ \mathrm{DIM}(A^{\circ} \cap \partial B) & \mathrm{DIM}(A^{\circ} \cap B^{\circ}) & \mathrm{DIM}(A^{\circ} \cap B^{-1}) \\ \mathrm{DIM}(A^{-1} \cap \partial B) & \mathrm{DIM}(A^{-1} \cap B^{\circ}) & \mathrm{DIM}(A^{-1} \cap B^{-1}) \end{bmatrix} \tag{5-8}
$$

郭仁忠（2000）则从地图学和实用GIS出发，提出对平面上拓扑关系按点-点、点-线、点-面、线-线、线-面、面-面目标对之间的相离、相邻、相交、包含和共位关系（含相邻、相交和包含），分为19种，并对相离、相邻、相交和严格包含关系作如下定义：

（1）相离（$A \parallel B$）：$A \cap B = \varnothing$；

（2）相邻（$A \mid B$）：$A \cap B = \partial A \cap \partial B = \neg \varnothing$；

（3）相交（$A < B$）：$A \subset B$；

（4）严格包含（$A \times B$）：$A^{\circ} \cap B^{\circ} = \neg \varnothing$。

Pigot S.（1992）将二维拓扑关系框架加以扩展，对多维欧氏空间实体的拓扑关系进行了描述。郭薇等（1997）提出了三维空间满足互斥、完备的空间最小集概念。陈军等（1998a，1998b，2002）在单纯形9I模型和 k-维空间实体形式化描述方法的基础上，用相离、相邻、相交、包含、相等、部分覆盖共6类基本关系对空间拓扑关系进行逻辑表达和语义描述，进而提出了三维空间的 k-单纯形（$0 \leqslant k \leqslant 3$）之间的拓扑关系的矩阵描述方法。李成名等（2000）则基于Voronoi图，研究提出了面条模型中动态空间关系推断的四邻近数序模型（4-adj模型），给出了惟一、完备性推断规则。吴立新等（2002）和齐安文等（2002）以三维地质体模拟为例，基于类三棱柱模型（ATP），定义和描述ATP的6个基本组成要素之间的三维空间拓扑关系。舒红等（1997）则对4I模型和9I交模型进行了时态扩展，提出了基于点集拓扑学的时态拓扑描述框架。Shi W. Z. 和 Guo W.（1999）提出不确定空间关系描述之模型。

§5.5 3D空间拓扑关系

3D GIS中的空间要素如结点、弧段、多边形、体等，都不是彼此孤立的，它们之间存在复杂的拓扑关系。通过存储某一要素相对于其他要素的位置关系，来建立较完善的拓扑库，可以使3D GIS具备一定的空间分析功能，并避免过于复杂的空间计算，提高3D GIS空间分析和查询的效率。因此，3D空间目标及其相互间拓扑关系形式化描述与表达是设计3D GIS空间数据库的重要基础，也是实现3D空间关系查询和空间分析的基本前提。

5.5.1 3D空间拓扑描述方法

3D GIS中的3D空间拓扑关系是2D空间拓扑关系的扩展，其拓扑完整性理论、拓扑关系形式化理论都是在2D GIS研究的基础上进一步发展的结果。其一般思路都是基于点集拓扑学，通过深入探讨空间关系的本质与判定机理，寻求空间关系形式化描述、表达与操作的方法与途径。3D空间拓扑关系的研究是3D空间构模研究的延续，不同

的空间构模方法对应不同的空间拓扑关系。迄今为止，国内外有关3D空间拓扑描述的主要研究成果如下。

1. 3D FDS拓扑描述

Molennar M.A.（1992）在原二维拓扑数据结构的基础上，定义了结点（Node）、弧（Arc）、边（Edge）和面（Face）四种几何元素之间的拓扑关系及其与点（Point）、线（Line）、面（Surface）和体（Solid）四种几何目标之间的拓扑关系，并显式地表达点和体、线和体、点和面、线和面间的is-in，is-on等拓扑关系，提出了一种基于3D矢量图的形式化数据结构（Formal Data Structure，FDS），如图5-27所示。Zeitouni等在此基础上，增加了垂直（vertical）及平行（relief）关系。

该模型存在3个问题：①只考虑了空间实体表面的划分和边界表达，没有考虑空间实体的内部结构，故仅适于表达诸如建筑物等形状规则的简单空间实体，难以表达地质及环境领域中没有规则边界的复杂3D空间实体（如矿体等）；②没有对3D空间实体及实体之间拓扑关系进行严格的定义及形式化描述，缺乏拓扑关系的完备性与独立性证明；③显式地存储is-in，is-on等拓扑关系，操作不便，影响系统效率。

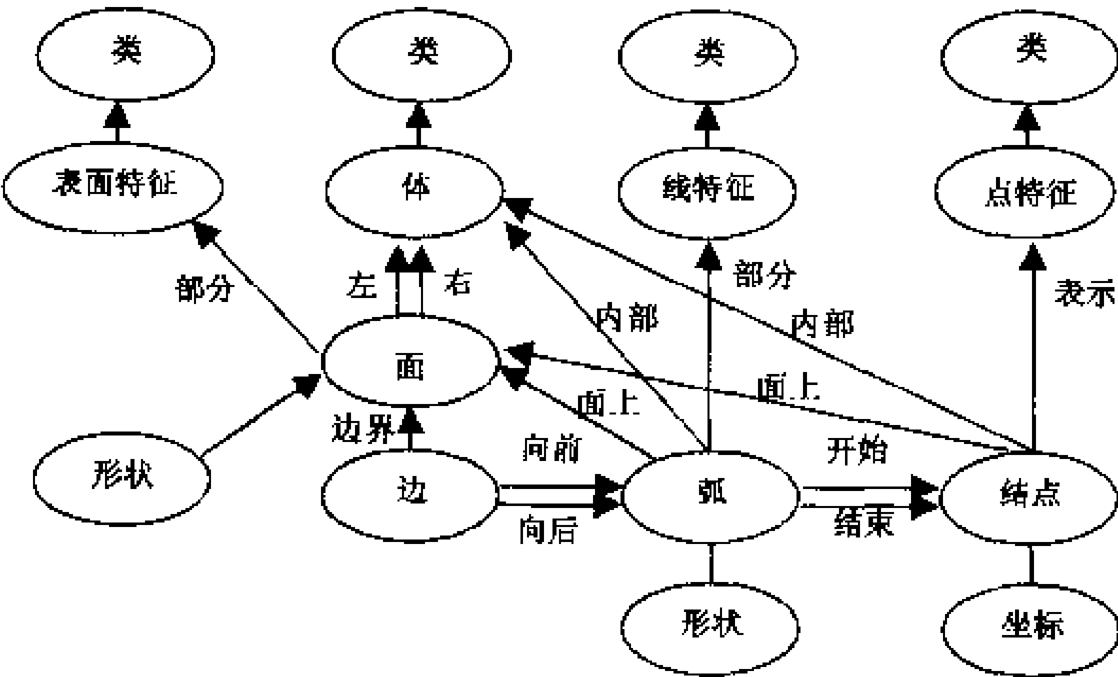


图5-27 3D FDS数据结构

（据Molenaar M. A.，1992）

2. 基于不规则四面体的3D拓扑描述

Pilouk M. et al.（1994）、Chen X. Y.（1995）等学者对基于不规则四面体的3D矢量数据模型进行了研究，将不规则四面体作为描述空间实体的基本体素，把任意一个3D空间实体剖分成一系列邻接但不重叠的不规则四面体，通过四面体间的邻接关系来反映空间实体间的某些拓扑关系。

该模型的优点是能够根据3D空间采样点的坐标值，有效地实现插值运算及图形的可视化功能，能快速进行几何和逻辑变换；缺点是仅考虑了空间实体内部结构的划分，没有考虑空间实体的表面形态，难以表达3D面状目标及线状目标；此外，随着数据精度的增加和数据量急剧增大，系统的查询速度受到影响。

3．基于“体划分”的3D矢量拓扑描述

李青元（1996）提出在3D空间中只有体划分是有限-互斥-完整划分的概念，不同的数据场类型对应不同的体划分方案。如图5-28所示：3D空间由不同的体充满整个研究区域，体与体之间由曲面片隔开；曲面片由一个外环和若干内环构成边界，并由若干条型值曲线定义曲面片的形状；相邻曲面片交于边，环由边有向连接成一个闭合回路；边由两个端结点限定边界，两个端点之间用一系列曲线型值点限定边的形状；不同的边交于结点。针对单一体划分下的矢量3D GIS模型，研究提出了以下5组拓扑关系：①结点-始边-终边；②边-起结点-终结点-环；③环-边-内邻曲面片；④曲面片-外边界环-内边界环-正面相邻体-负面相邻体；⑤多面体-曲面片。基于“一片二层”，并通过“界面引入-体划分”，讨论了5组拓扑关系的动态地建立与维护方法，即首先对研究区从无结构的3D空体建立拓扑关系，然后不断引进分界面（曲面片）进行3D体划分，引进一个界面就将原来的一个多面体一分为二，并调整拓扑关系，直至达到所要求的体划分。该模型的缺点是程序实现非常困难，人机交互工作量巨大，应用难度大。

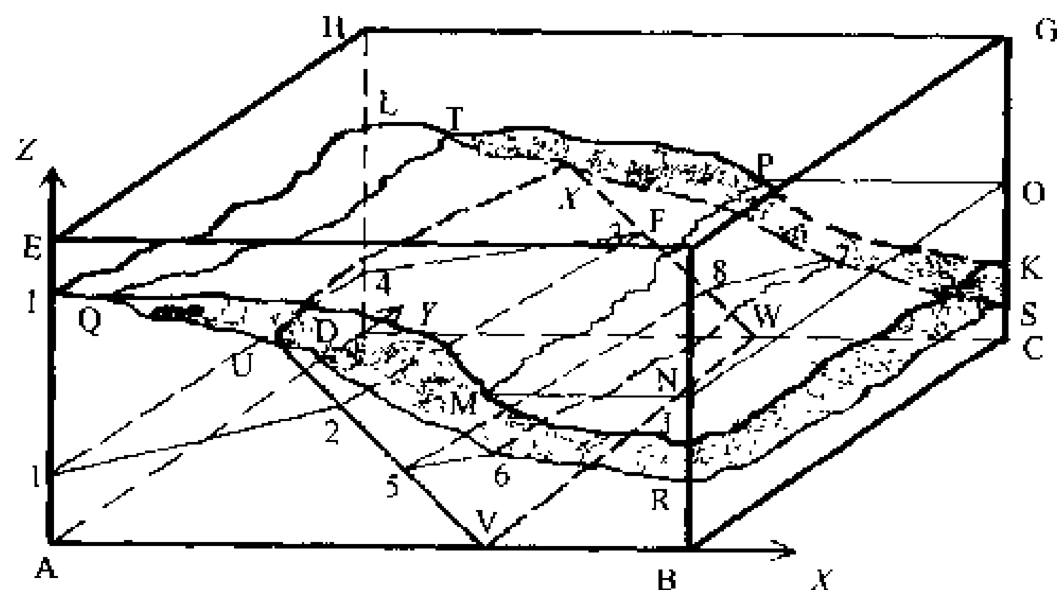


图5-28　3D矢量结构GIS拓扑关系

（李青元，1996）

4．顾及空间剖分的3D拓扑描述

舒红（1997）、陈军（1998）、郭薇（1999）等提出3D空间实体可分成点实体（Point entity）、线实体（Line entity）、面实体（Surface entity）和体实体（Body entity），如图5-29所示。任意一个3D空间实体是一个可定向的 n -维伪流形（n-pseudo-manifold），（$0\leqslant n\leqslant 3$），它对应于一个具有良好单纯形结构的 n -单纯复形（n-complex），在几何上可剖分成若干个维数小于或等于它的、连通但不相互重叠的 k -单纯形（k-simplex），（$k<=$n）。0-单纯形（0-simplex）、1-单纯形（1-simplex）、2-单纯形（2-simplex）、3-单纯形（3-simplex）和平面（face）是组成3D空间实体的5种基本几何元素（Geometric elements），其中平面是所有具有相同法线矢量的2-单纯形的集合，用于减少数据冗余及更有效地表达具有规则边界的面实体和体实体。

该拓扑模型兼顾了矢量数据模型和镶嵌数据模型的优点，能较好地表达3D实体及

其相互间的空间拓扑关系。

5. 基于矢栅混合的3D拓扑描述

李清泉等（1998）在3D栅格矢量混合数据结构中，以八叉树结构作整体描述，以3D拓扑矢量边界表示作局部描述。对于一个不规则矿体，主要由3步来表示：① 3D实体的最小边界长方体化表示（MBC），即用一个长方体套管来完全套上所要表示的实体形成一个粗略表示，该套管由实体的最小边界坐标和最大边界坐标确定；② 3D实体的多个MBC表示，即在第一步粗略表示之后，进一步细分该长方体，以多个（一般为8个）相联系的MBC来表示；③ 3D实体的拓扑矢量边界表示（TVBR），即对实体经过上述两步表示后，再考虑进拓扑关系，即结点-边、边-结点-环、环-边-曲面片、曲面片-环-体、体-体以及点或线目标在面内或体内的关系，对3D实体进行拓扑表达。

程朋根等（2001）采用面向对象的技术，提出了矢栅集成的面向对象混合数据模型，以复杂体-体-面-线-点对象之间的逻辑关系来建立对象之间的拓扑关系，并设计了部分对象的数据结构，以铁矿床为例进行了数据结构描述。

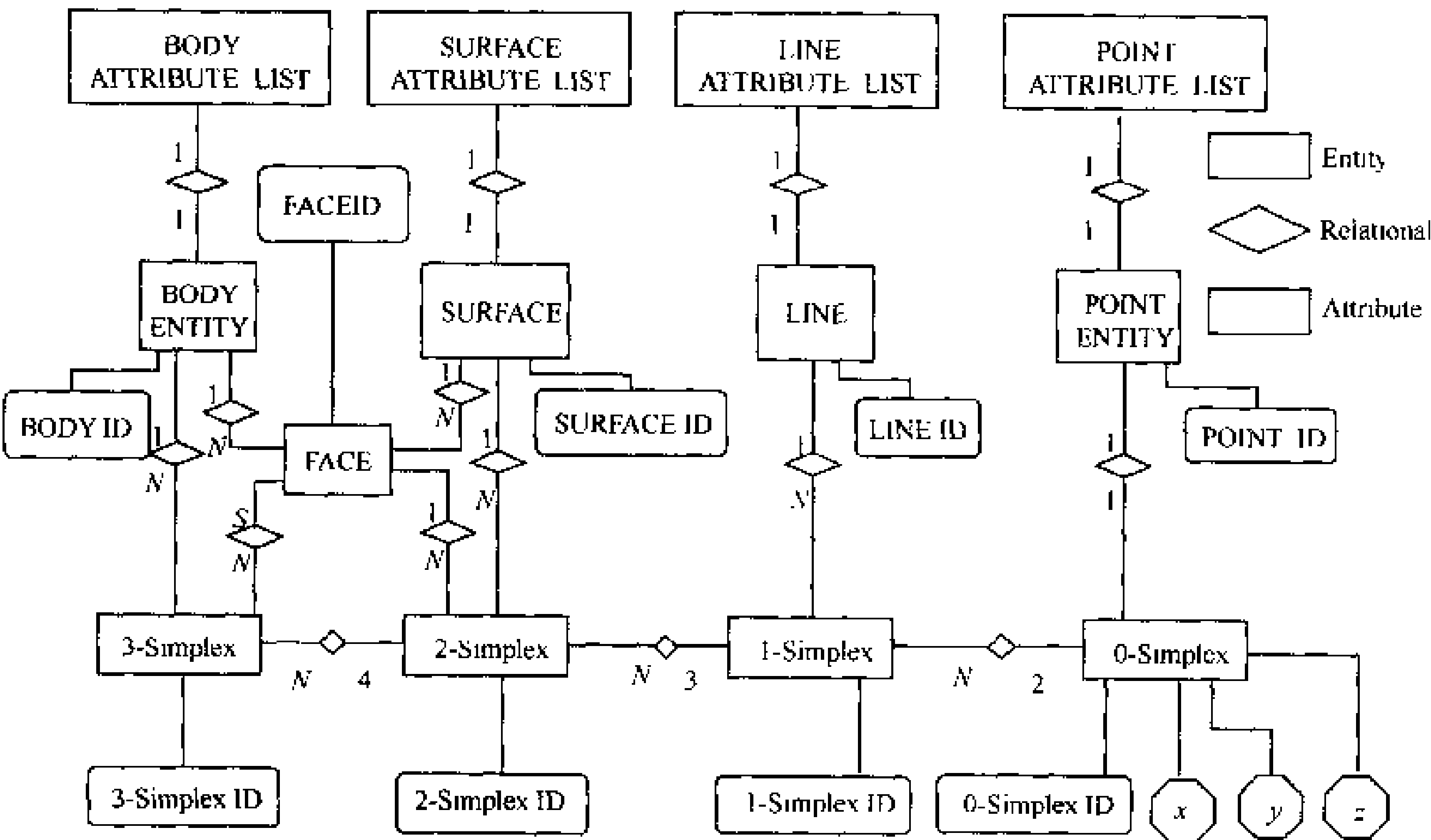


图 5-29 顾及空间剖分的3D拓扑描述

（陈军，1998）

以上模型有利于3D图形的显示，还能精确表达空间体之间的拓扑关系，但在软件实现上较困难。

6. 面向对象的3D矢量GIS拓扑描述

孙敏等（1998）提出面向对象的3D GIS空间矢量网状数据模型，以链表作为基本结构，把点、线、面、体看作是3D GIS的基本元素，以每个元素为对象设计数据结

构。对每一个基本元素定义一个类，拓扑关系包含在类的成员变量及成员函数中。程序设计中只针对每个元素，运用其成员变量，编写其成员函数，程序通过调用其成员函数即可实现拓扑关系的自动建立。用 Visual C++ 可写出 3D 矢量 GIS 数据模型如下：

class CBasePoint：定义基础点元素类；class CorPoint：CbasePoint：定义坐标点元素类，由基础点元素派生；class CnodePoint：CbasePoint：定义结点元素类，由基础点元素派生；class CBaseLine：定义基础线元素类；class CnodeLine：CbaseLine：定义结线元素类，由基础线类派生；class Cpoly：定义面元素类；class Cform：定义体元素类。

设计表结构来保存指针，通过表结构中的这种指针来表达空间拓扑关系。基本拓扑关系如下：点-线关系（点为线结点，线由点构成）；点-面关系（面由点通过算法生成，面包含点）；线-面关系（面由线拟合生成，线为面的结线）；面-体关系（体由面构成，面为多体公用面）。隐含的关系有：线-线关系（多线共用结点即相交）；面-面关系（多面共用结线即相交）；体-体关系（多体相邻或包含）。上述拓扑关系基本上能全面反映已有的空间拓扑关系，由于以点、线、面、体基本对象进行 3D 构模，故适合面向对象的空间操作与分析。

7. 基于“层次树”的 3D 矢量拓扑描述

陈云浩等（1999）从矿业应用实际出发，提出了采用“层次树”3D GIS 矢量拓扑结构，用 6 组关系来描述矿山 GIS 的 3D 矢量结构的空间拓扑关系。如图 5-30 所示：①复杂地物-体关系：即复杂地物与组成它的体域的关系。在该拓扑关系中加入对复杂地物属性的描述或指向属性记录文件的指针；②体-复杂地物-曲面关系：即体域与由其所构成的复杂地物、体域与包围该体域的曲面以及与该体域相邻的体域的关系。体拓扑结构中可加入对体域属性的描述；③曲面-环-体域关系：即曲面与组成曲面的环以及该曲面所包围的体域（正面邻体）和包围该曲面的外部体域（负面邻体）的关系。一个曲面可能由若干个环构成，其外环取正，内环取负值。在曲面拓扑结构中加上若干条弧（如等高弧）或插值函数，就可确定该曲面的空间形态；④环-弧-曲面关系：即环与构成该环的弧以及该环所包围的内部曲面（内邻曲面）和包围该环的外部曲面（外邻曲面）的关系；⑤弧-结点-环的关系：即弧与该弧的起结点、终结点及包含该弧的环的关系。环有正负之分，环方向与弧的方向一致时，取正号；反之取负。在弧拓扑结构中加一系列坐标串，可确定该弧的形态；⑥结点-弧的关系：即结点及从该结点出发的离开弧段和

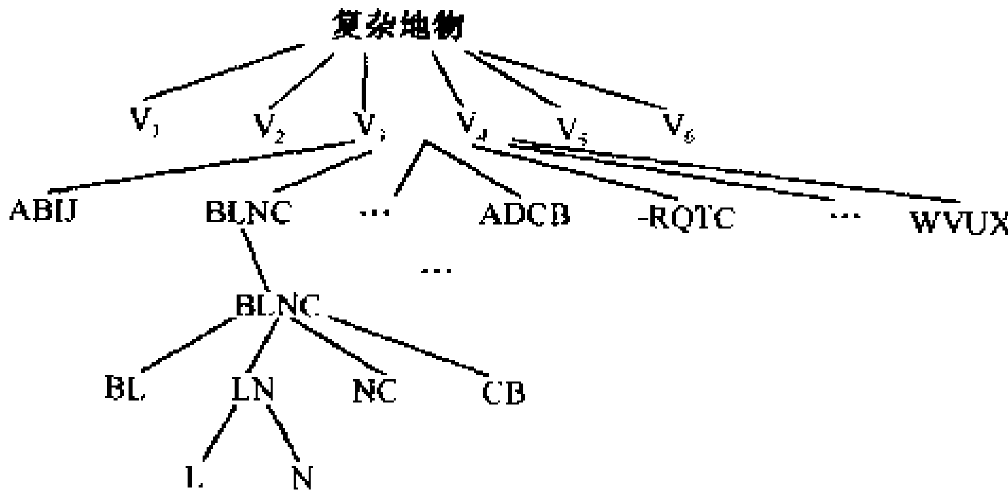


图 5-30 基于“层次树”的 3D 矢量拓扑结构
（陈云浩，1999）

以该结点为终点的到达弧段的关系。

该模型描述的中心是体域，而把面域、线域、点域看作特殊的体域，并按体域形式存贮，以点、线、面、体为基本对象进行构模，故适合于面向对象的空间操作与分析；所建立的拓扑关系明确，层次分明，能降低软件算法的实现难度。

8. 基于岩土建模的3D拓扑描述

王笑海（1999）研究了基于3D拓扑格网结构的GIS地层模型，讨论了该模型的拓扑问题；陈健（2001）研究了用多层DEM模型来建立3D数字地层模型，并建立了局域拓扑模型。这些研究主要依托岩土工程问题，不太适合具有深部特征和钻孔偏斜的地质、采矿问题；而且格网模型和DEM模型都是基于面构模的，不能实现矿体内部的模型构建和拓扑表达。

5.5.2 基于GTP的真3D拓扑

基于GTP的拓扑关系描述是一个复杂的过程。GTP的6个基本组成元素之间可以构造出6组共36种拓扑关系，以点组为例：点-点、点-TIN边、点-侧边、点-TIN面、点-侧面和点-GTP共6种。上述关系可以按多种方法、多种表格形式进行设计，其关键是效率与数据冗余的合理平衡。表格太少或过于简单，虽然数据冗余少，但检索查询效率低；反之，表格过多或过于复杂，则数据冗余大，但检索查询效率高。

1. 拓扑关系表设计原则

选用拓扑关系和设计拓扑关系表时，综合考虑了以下原则：①完整性原则：即要能够全面反映空间6类基本元素之间的所有36种拓扑联接关系；②少冗余原则：为加快查找速度，对于经常使用的关系，允许保留一定的冗余；但应在保证系统一致性的前提下尽量减少数据冗余；③便捷性原则：为方便查找与使用，尽量减少表间链接追索的次数，以加快查询速度，即最多检索两张表就能得到36种拓扑关系中的任何一种。

2. 拓扑关系表结构设计

基于以上原则，选择并设计了6张表来描述GTP内部各要素之间的拓扑关系，即：①TIN edge-Node-TIN face；② GTP-face（TIN face and Side face）-Node；③Side edge-Node-Side face-TIN face；④ TIN face-TIN edge-Side face-Upper GTP-Lower GTP；⑤Side face-Edge（TIN edge and side edge）-GTP；⑥Node-TIN edge-Side edge-GTP。通过这6张表，可以便捷地产生其中未包含的拓扑关系。比如，Node-TIN face的关系可以从⑥Node-TIN edge-Side edge-GTP表和①TIN edge-Node-TIN face表中得到，而GTP-TIN edge关系则可以从②GTP-face（TIN face and side face）-node表和⑤Side face-edge（TIN edge and side edge）-GTP中得到。

图5-31所示为3个GTP单元的一种典型组合，上述6张拓扑关系表的具体内容见表5-6～表5-11（Wu L.X. et al.，2003）。

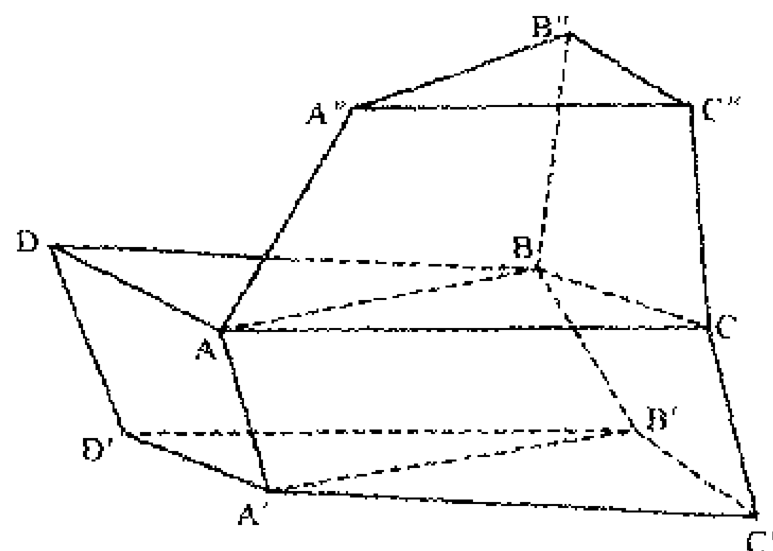


图 5-31 基于 GTP 的拓扑关系示意图
(Wu L.X. et al., 2003)

表 5-6 TIN-edge, Node 和 TIN-face 之间的拓扑关系表

TIN-edge	Node	TIN-face
AB	A, B	ABD, ABC
AC	A, C	ABC
AD	A, D	ABD
BC	B, C	ABC
BD	B, D	ABD
A′D′	A′, D′	A′B′D′
A′C′	A′, C′	A′B′C′
A′B′	A′, B′	A′B′D′, A′B′C′
B′C′	B′, C′	A′B′C′
B′D′	B′, D′	A′B′D′
A″B″	A″, B″	A″B″C″
A″C″	A″, C″	A″B″C″
B″C″	B″, C″	A″B″C″

表 5-7 GTP, Face (TIN-face and side-face) 和 Node 之间的拓扑关系表

GTP	Face (TIN-face and side-face)	Node
ABDA′B′D′	ABD, A′B′D′, BDD′B′, ABB′A′, ADD′A′	A, B, D, A′, B′, D′
ABCA′B′C′	ABC, A′B′C′, BCC′B′, ABB′A′, ACC′A′	A, B, C, A′, B′, C′
ABCA″B″C″	ABC, A″B″C″, ACC″A″, ABB″A″, BCC″B″	A, B, C, A″, B″, C″

表 5-8 Side-edge, Node, Side-face 和 TIN-face 之间的拓扑关系表

Side-edge	Node	Side-face	TIN-face
AA′	A, A′	ADD′A′, ACC′A′, ABB′A′	ABC, ABD, A′B′D′, A′B′C′
BB′	B, B′	BDD′B′, ABB′A′, BCC′B′	ABC, ABD, A′B′D′, A′B′C′
CC′	C, C′	ACC′A′, BCC′B′	ABC, A′B′C′
DD′	D, D′	ADD′A′, BDD′B′	ABD, A′B′D′
AA″	A, A″	ABB″A″, ACC″A″	ABC, ABD, A″B″C″
BB″	B, B″	ABB″A″, BCC″B″	ABC, ABD, A″B″C″
CC″	C, C″	BCC″B″, ACC″A″	ABC, A″B″C″

表 5-9 TIN-face, TIN-edge, Side-face 、Upper-GTP 和 Lower GTP 之间的拓扑关系表

TIN-face	TIN-edge	Side-face	Upper GTP	Lower GTP
ABC	AB, BC, AC	ABB′A′, ACC′A′, BCC′B′, ABB″A″, BCC″B″	ABCA″B″C″	ABCA′B′C′
ABD	AB, BD, AD	ABB′A′, ADD′A′, BDD′B′, ABB″A″		ABDA′B′D′
A′B′C′	A′B′, B′C′, A′C′	ABB′A′, ACC′A′, BCC′B′	ABCA′B′C′	
A′B′D′	A′B′, B′D′, A′D′	ABB′A′, ADD′A′, BDD′B′	ABDA′B′D′	
A″B″C″	A″B″, B″C″, A″C″	ABB″A″, ACC″A″, BCC″B″		ABCA″B″C″

表 5-10 Side-face, Edge (TIN-edge and side-edge) 和 GTP 之间的拓扑关系表(Wu L.X. et al., 2003)

Side-face	Edge (side-edge and TIN-edge)	GTP
ABB′A′	AB, BB′, A′B′, AA′	ABDA′B′D′, ABCA′B′C′
ACC′A′	AC, CC′, A′C′, AA′	ABCA′B′C′
ADD′A′	AD, DD′, A′D′, AA′	ABDA′B′D′
BDD′B′	BD, BB′, B′D′, DD′	ABDA′B′D′
BCC′B′	BC, CC′, B′C′, BB′	ABCA′B′C′
AA″B″B	A″B″, BB″, AB, AA″	ABCA″B″C″
AC C″A″	A″C″, C C″, AC, AA″	ABCA″B″C″
BCC″ B″	B″C″, C C″, BC, BB″	ABCA″B″C″

表 5-11 Node, TIN-edge, Side-edge 和 GTP 之间的拓扑关系表

Node	TIN-edge	Side-edge	GTP
A	AB, AC, AD	AA′, AA″	ABCA′B′C′, ABCA″B″C″, ABDA′B′D′
B	AB, BC, BD	BB′, BB″	ABCA′B′C′, ABCA″B″C″, ABDA′B′D′
C	AC, BC	CC′, CC″	ABCA′B′C′, ABCA″B″C″
D	AD, BD	DD′	ABDA′B′D′
A′	A′B′, A′C′, A′D′	A A′	ABCA′B′C′, ABDA′B′D′
B′	A′B′, B′C′, B′D′	BB′	ABCA′B′C′, ABDA′B′D′
C′	A′C′, B′C′	CC′	ABCA′B′C′
D′	A′D′, B′D′	DD′	ABDA′B′D′
A″	A″B″, A″C″	AA″	ABCA″B″C″
B″	A″B″, B″C″	BB″	ABCA″B″C″
C″	A″C″, B″C″	CC″	ABCA″B″C″

3. 拓扑关系推导

某些扩展的拓扑关系也可以从这 6 张表格中推导出来。例如：GTP-GTP 关系可以从 GTP-face (TIN face and side face) -node 表中隐式推导出来：①如果两个 GTP 单元共享某一 TIN 面，则该两个 GTP 单元为 TIN 面相邻，属于不同的地层；②如果两个 GTP 单元共享某一侧面，则该两个 GTP 单元为侧面相邻，属于同一地层；③ 如果两个 GTP 单元仅仅共享某一结点，则该两个 GTP 单元为结点相邻，属于上下不同地层；④如果两个 GTP 单元共享同一界面的两个结点，则该两个 GTP 单元为 TIN 边相邻，属于上下不同地层；⑤如果两个 GTP 单元共享不同界面的两个结点，则该两个 GTP 单元为侧边相邻，属于同一地层。

此外，如表 5-12 (Wu L.X. et al., 2003) 所示，GTP 单元 6 要素之间的包含、邻

接和相离关系也可以从上述6张表中推导出来。

表 5-12 GTP 各要素之间的3类基本拓扑关系

Relation / Primitives	Inclusion	Adjacent	Disjoint
Node (P_1)	P_2, P_3, P_4, P_5, P_6	P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6
TIN-edge (P_2)	P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6
side-edge (P_3)	P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6
TIN-face (P_4)	P_6	P_1, P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6
side-face (P_5)	P_6	P_1, P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6
GTP (P_6)	—	P_1, P_2, P_3, P_4, P_5, P_6	P_1, P_2, P_3, P_4, P_5, P_6

4. 拓扑关系查询与检查

基于以上拓扑描述，可以建立GTP的拓扑维护与查询系统，并实现基于这些库的拓扑查询及GTP增加、删除和修改等功能。

拓扑关联的查询和检查基本原理为：首先指定任一个有标识号的GTP作为初始体，然后通过与之相邻的点、线或面找到与之相邻的GTP，凡是通过侧面或棱边相邻的标记为1，通过顶点、TIN面或TIN边相邻的标记为2，都不相邻的标记为0。标记为1的GTP为同一层，标记为2的为相邻层，标记为0的为拓扑相离。对于某一地层，任选其中的一个GTP搜索该层中的所有其他GTP，若均标记为1，则拓扑描述正确；否则，该地层可能出现了分叉现象，或受到了断层切割，或拓扑描述不正确。对于某一地层，任选其中的一个GTP搜索该层的上邻层或下邻层中的所有GTP，若均标记为2，则拓扑描述正确；否则，该地层可能出现了尖灭现象，或受到了断层切割，或拓扑描述不正确。

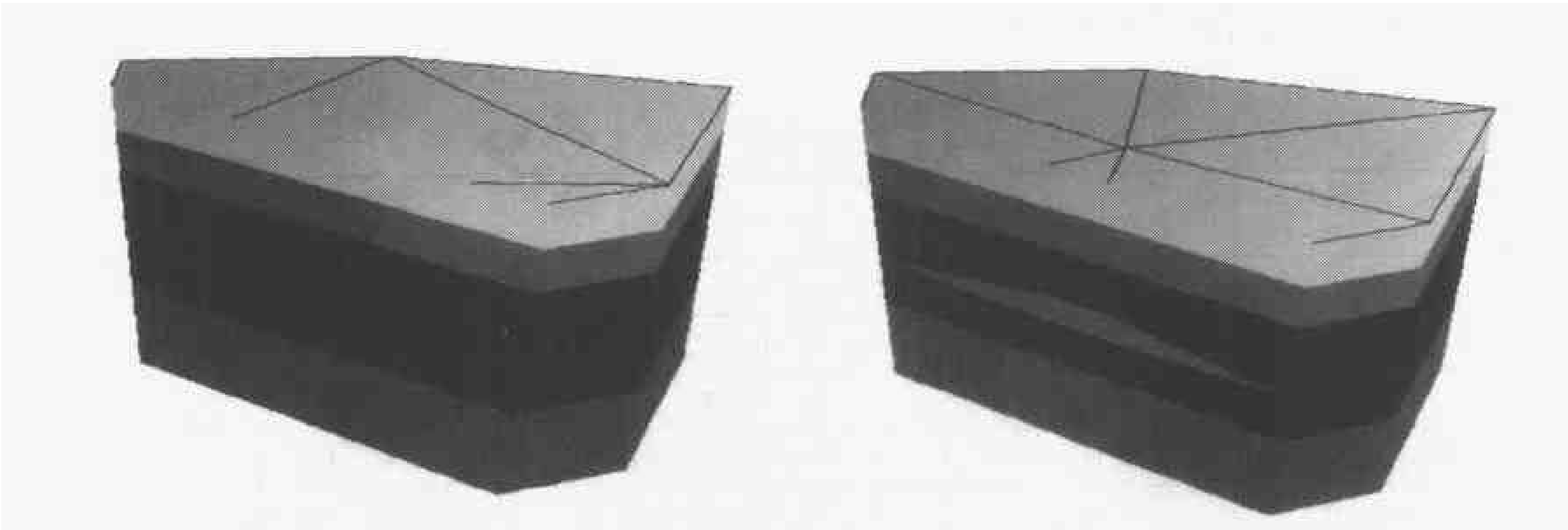
5. 拓扑关系维护

由于GTP模型的数据结构中包含了各元素之间的拓扑关系，所以在进行TIN与GTP局部重构时，GTP的拓扑关系也需进行动态维护。

以新的钻孔点加入为例：首先根据该点惟一的标识号将其加入到点库中；再由TIN的生成程序按照Delaunay法则局部重构TIN；然后通过TIN的数据结构，找到组成TIN边的该点。依次类推，直到由新点参与构成新的GTP。若有新的地层加入时（如图5-32所示），原来的地层在局部形成了新的地层界面，新地层将原来地层重新进行了分割。切割后形成新的GTP，在局部对原来的GTP进行拓扑重构。

若有新的空间点加入（如图4-21所示），新的三角形将原来的GTP切分成两个GTP，这样在两个新的GTP之间就通过新的三角形建立邻接关系，成为公共TIN面。而两个新的GTP与原来的GTP之间也可以通过原来的上、下TIN面建立邻接关系，上、下TIN面成为公共面。而新的GTP通过侧面与周围和它相邻的GTP建立邻接关系。

以上操作都是只局部重新生成并修改了拓扑关系的表达，而不需重新生成并修改整



(a) 原始模型　　(b) 新地层加入后的更新模型

图 5-32　新地层加入后拓扑维护实验结果

体的拓扑关系，这就使得 GTP 局部拓扑维护与动态更新效率较高。

§5.6　其他空间关系

Florence J.（1996）提出了空间相离（disjoint）关系，认为空间相离关系在空间关系中占有较大比例；陈军等（1997，1999）总结了相离（disjoint）、相接（touch）、相交（cross）、包含于（in）和交叠（overlap）共 5 种基本拓扑关系后指出：空间目标之间还存在靠近（close，near-to）等模糊空间关系、穿越与进入等反映目标运动状态的空间关系及目标之间的不确定性空间关系。王家耀（2001）还定义了相等（equal）、包含（contain）、覆盖（cover）、被覆盖（coverby）等 4 种基本空间关系。此外，人们还提出了空间相似关系和空间相关关系的概念。

5.6.1　3D 空间拓扑关系全集

作者研究认为，可以采用相离（disjoint）、相等（equal）、相接（touch）、相交（cross）、包含于（in）、包含（contain）、交叠（overlap）、覆盖（cover）、被覆盖（coveredBy）、进入（enter）、穿越（pass）和被穿越（passBy）等共 12 种基本空间关系表达 3D 空间中的以下 10 类有理论价值和实际意义的空间拓扑关系。

（1）点-点空间关系 2 种



(a) 相离　　(b) 相等

图 5-33　点-点空间拓扑关系

(2) 点-线空间关系 3 种

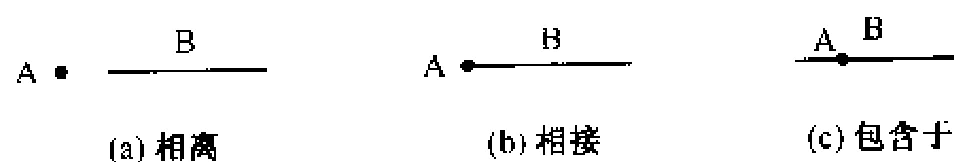


图 5-34 点-线空间拓扑关系

(3) 点-面空间关系 3 种

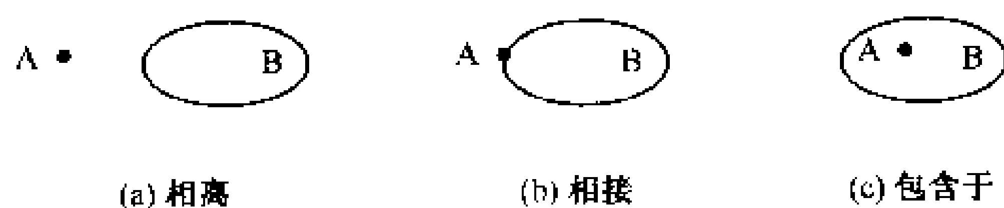


图 5-35 点-面空间拓扑关系

(4) 点-体空间关系 3 种

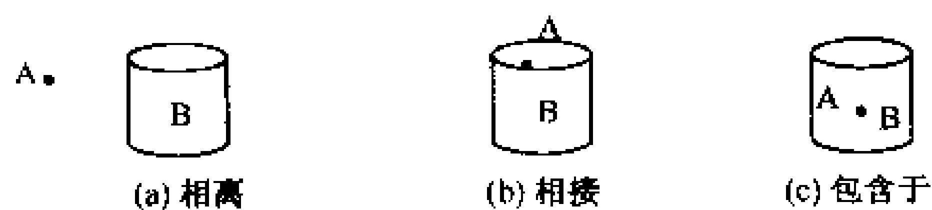


图 5-36 点-体空间拓扑关系

(5) 线-线空间关系 7 种

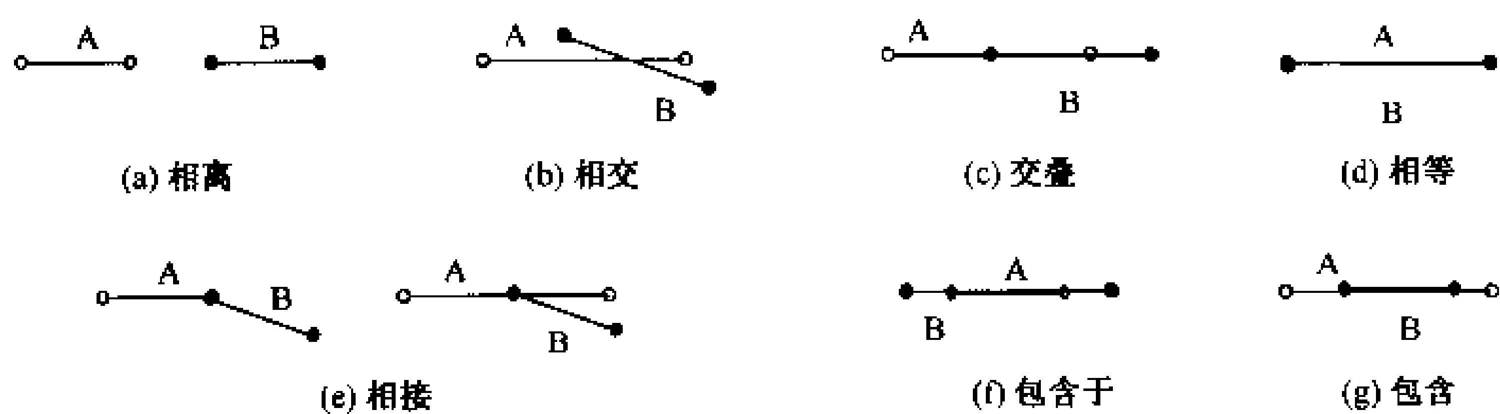


图 5-37 线-线空间拓扑关系

(6) 线-面空间关系 5 种

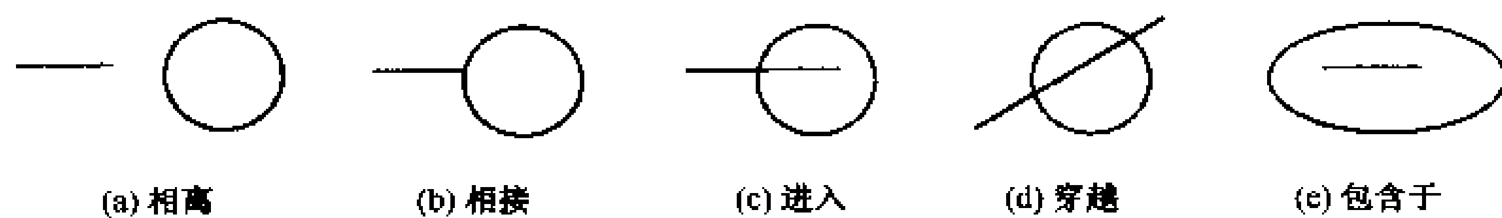


图 5-38 线-面空间拓扑关系

(7) 线-体空间关系 5 种

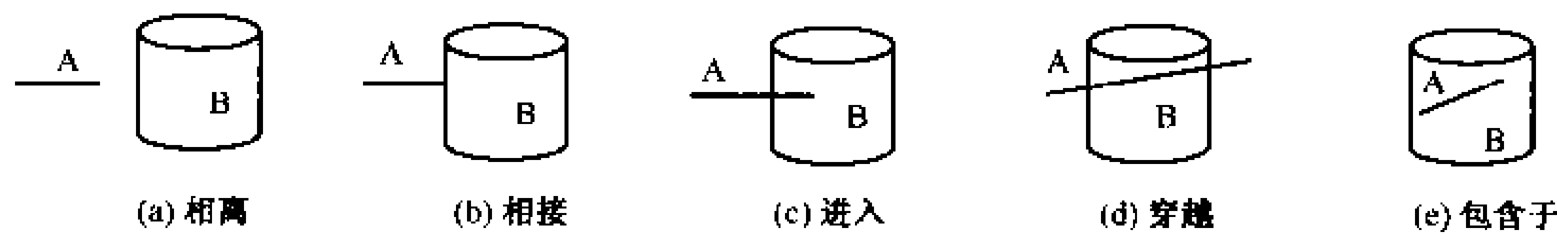


图 5-39 线-体空间拓扑关系

(8) 面-面空间关系 10 种

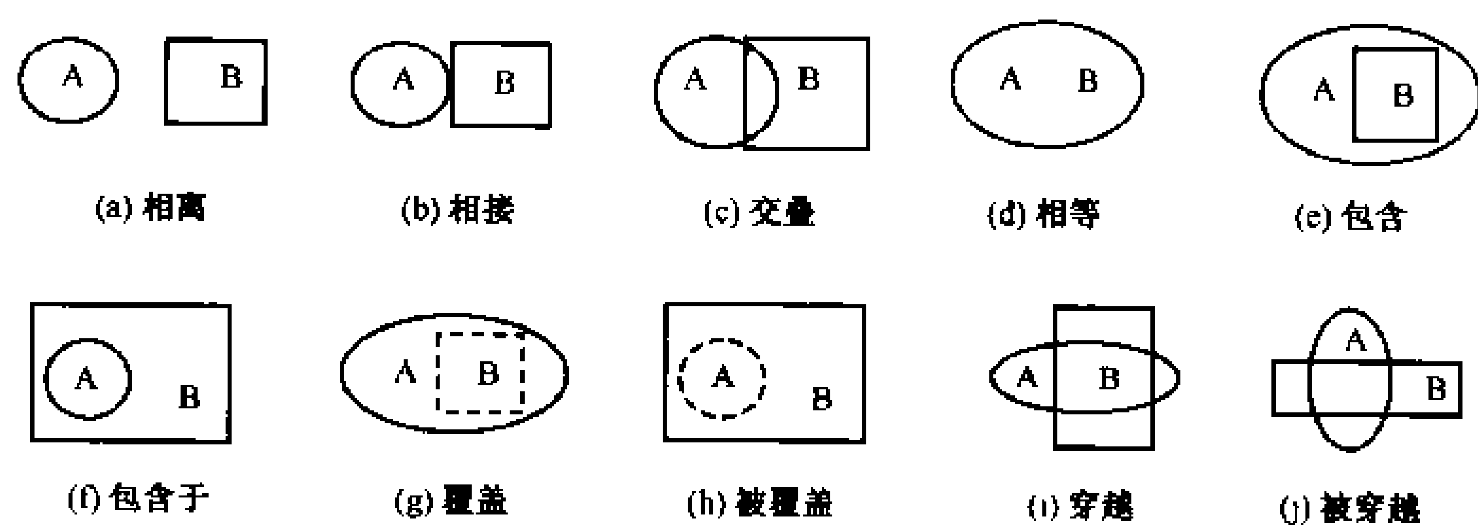


图 5-40 面-面空间拓扑关系

(9) 面-体空间关系 8 种

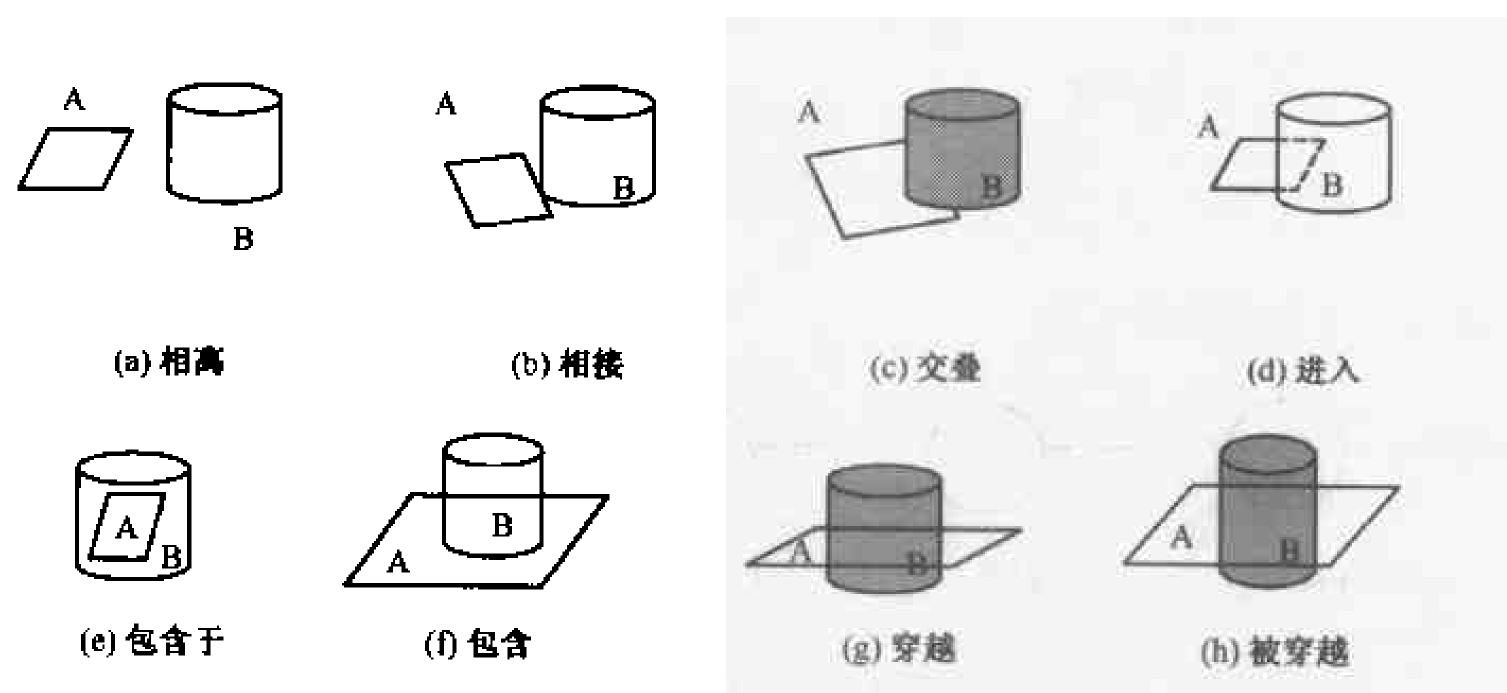


图 5-41 面-体空间拓扑关系

（10）体-体空间关系 8 种

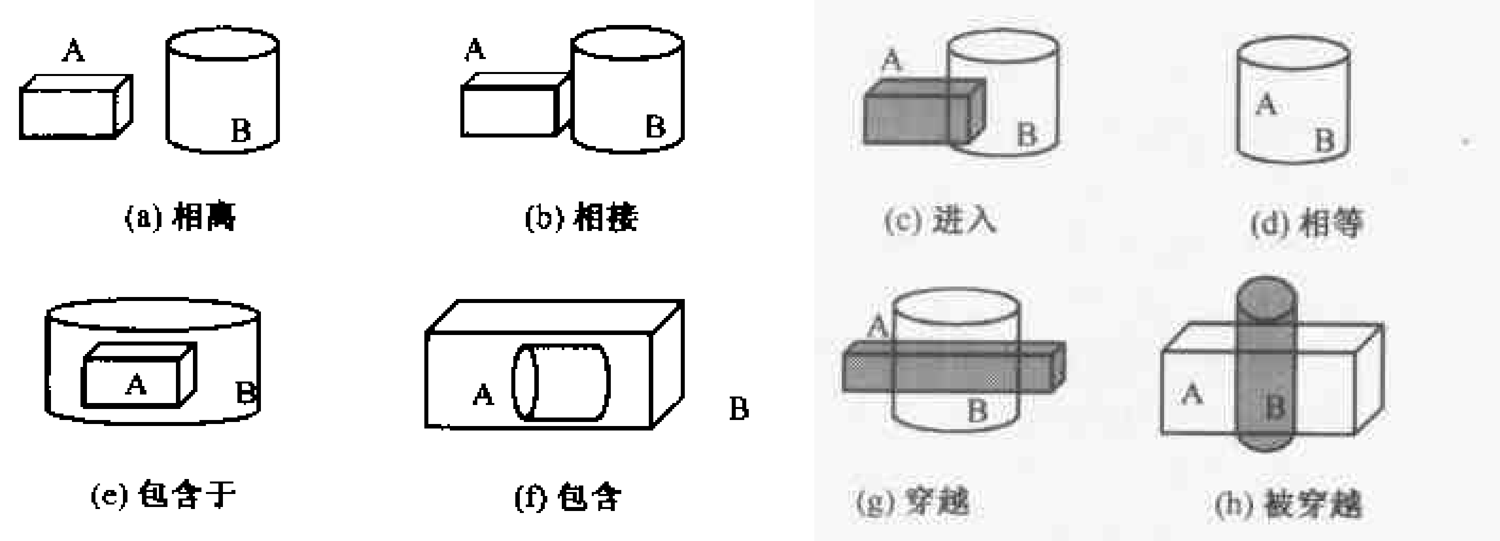


图 5-42　体-体空间拓扑关系

以上 10 类 54 种空间拓扑关系可以基于拓扑学理论，进行适当定义和数学描述。其中 2D 拓扑关系的有关数学描述可参见王家耀（2001），3D 拓扑关系的有关数学描述可参见 Silvio L.（1997）。各类拓扑关系的应用需根据实际情况具体分析，如相接关系有单点相接、两点相接、多点相接和线相接、面相接等多种情况；穿越与被穿越关系在城市 GIS、矿山 GIS 中有重要意义。

5.6.2　空间相似与空间相关关系

所谓空间相似关系，是指空间目标几何形态上的相似或空间目标（群）结构上的相似。基于几何形态相似的经典应用有：德国地球物理学家魏格纳根据大西洋两岸大陆轮廓线的相似吻合特征，经分析提出了著名的大陆漂移学说；扫描地图的地物要素识别与卫星影像中几何特征自动提取，也频繁使用几何形态相似的原理。基于形态相似的分析有两条途径（郭仁忠，2000）：①相似变换下的图形吻合度分析；②基于图形形态参数的聚类分析或相关分析。基于结构相似的典型应用为遥感影像特征识别与特征提取，包括基于树状结构、扇状结构和网状结构的河网识别，基于棋盘状和放射状结构的路网识别，基于集块状、散列状结构的居民地的识别等。结构相似同时还是智能化空间查询（基于空间条件、属性条件及其综合）的一个重要研究内容。

所谓空间相关关系，是指两个地理目标之间存在的某种互为因果的关系，如河流溢洪与大气降雨、植被覆盖、地表径流等密切相关，植被覆盖与地理纬度、海拔高度、坡度坡向、气候条件等密切相关。空间相关分析的数学基础是统计学和矩阵理论。

作者认为，前述的空间拓扑关系不同于此处的空间相似关系及空间相关关系。前者属于空间目标的基础空间关系，与空间目标的几何形态密切相关，且直接依赖于关系表来进行描述和表达；而空间相似关系则属于空间目标的外延空间关系，虽然也与空间目标的几何形态密切相关，但独立于数据组织和拓扑结构；空间相关关系则为空间现象或空间过程的空间因果关系，也独立于数据组织和拓扑结构，可以通过数据挖掘获取（详见第二十三章）。

参 考 文 献

曹　菡,陈　军,杜道生. 2001. 空间目标方向关系的定性扩展描述. 测绘学报. 30(2): 162～167

陈　健. 2001. 三维地层空间信息系统结合三维有限元的研究与实现[博士学位论文]. 中国科学院武汉岩土研究所

陈　军,崔秉良. 1997. 用 Voronoi 方法为 Mapinfo 扩展拓扑功能. 武汉测绘科技大学学报. 22(3): 195～200

陈　军,郭　薇. 1998a. 基于剖分的三维拓扑 ER 模型研究. 测绘学报. 27(4): 308～316

陈　军,郭　薇. 1998b. 三维空间实体间拓扑关系的矩阵描述. 武汉测绘科技大学学报. 23(4): 359～363

陈　军,赵仁亮. 1999. GIS 空间关系的基本问题与研究进展. 测绘学报. 28(2): 95～102

陈　军. 2002. Voronoi 动态空间数据模型. 测绘出版社.

陈云浩,郭达志. 1999. 一种 3DGIS 数据结构的研究. 测绘学报, 2(1): 41～44

程朋根,龚健雅. 2001. 地勘工程 3D 空间数据模型及其数据结构设计. 测绘学报. 30(1): 74～81

陈require金. 1998. 矿井导线网空间关系和操作的半序集模型. 测绘学报. 27(1): 58～62

郭仁忠. 2000. 空间分析. 武汉:武汉测绘科技大学出版社. 153～220

郭　薇,陈　军. 1997. 基于点集拓扑学的三维拓扑空间关系形式化描述. 测绘学报. 26(2): 122～127

郭　薇. 1998. 顾及空间剖分的三维拓扑空间数据模型[博士学位论文]. 武汉测绘科技大学

郭　薇,詹　平,郭　箐. 1999. 面向地理信息系统的 3D 空间数据模型. 江西科学, 17(2): 77～82

侯恩科,吴立新. 2002. 面向地质建模的三维体元拓扑数据模型研究. 武汉大学学报(信息科学版), 27(5): 467～472

胡　勇,陈　军. 1997. 基于 Voronoi 图的空间邻近关系表达和查询操作. 中国 GIS 协会第二届年会论文集. 346～356

李成名. 1998. 基于 Voronoi 图的空间关系描述、表达与推断[博士学位论文]. 武汉测绘科技大学

李成名,陈　军. 2000. 面条模型中空间关系动态推断方法研究. 遥感信息. (2): 14～17

李清泉. 1998. 基于混合结构的 3DGIS 数据模型与空间分析研究[博士学位论文]. 武汉测绘科技大学

李清泉,李德仁. 1998. 3D 空间数据模型集成的概念框架研究. 测绘学报, 27(4): 325～330

李青元. 1996. 3D 矢量结构 GIS 拓扑关系研究[博士学位论文]. 中国矿业大学(北京校区)

李青元. 1997. 3D 矢量结构 GIS 拓扑关系及其动态建立. 测绘学报, 26(3): 235～240

刘亚彬,刘大有. 2001. 地理信息系统中空间对象间拓扑关系推理. 软件学报. 12(12):1059～1063

刘文宝,邓　敏,夏宗国. 2001. 论 GIS 中的模糊动态空间关系. 煤炭学报. 26(4): 374～379

毛建华,何　挺,刘春燕. 2000. 空间关系符号表示及其推理. 江西师范大学学报(自然科学版). 24(4): 367～369

齐安文等. 2002. 一种新的三维地学空间构模方法——类三棱柱法. 煤炭学报, 27(2): 158～163

舒　红等. 1997. 时空拓扑关系定义及时态拓扑关系描述. 测绘学报. 26(4): 299～306

孙　敏,唐小明,赵仁亮. 1998. 面向对象的 3D 矢量 GIS 数据模型及拓扑关系的建立. 测绘通报, (7): 11～14

孙玉国. 1993. 拓扑关系描述语 2D T-string 空间关系表达[博士学位论文]. 武汉测绘科技大学.

王家耀. 2001. 空间信息系统原理. 科学出版社. 220～242

王笑海. 1999. 基于 3D 拓扑格网结构的 GIS 地层模型研究[博士学位论文]. 中国科学院武汉岩土力学研究所

吴立新等. 2002. 三维地学模拟与虚拟矿山系统. 测绘学报, 31(1): 28～33

邬　伦等. 2001. 地理信息系统——原理、方法与应用. 北京:科学出版社. 61～64

闫浩文. 2001. 空间方向关系的概念、计算和形式化描述模型研究[博士学位论文]. 武汉测绘科技大学

赵仁亮等. 2000. 基于 V9I 的空间关系映射与操作. 武汉测绘科技大学学报, 25(4):318～323

Abdelmoty A. I. and Williams M. H., 1994. Approaches to the representation of qualitative spatial relationship for geographic database. Geodesy. 40: 204～216

Ai T. H., 1998. A topological relation description for spatial objects with uncertainity boundaries. In: proc. SIST'98. 394～398

Chang S. K., Shi Q. S. and Yan C. W., 1987. Iconic indexing by 2D-strings. IEEE Transactions on Pattern Analysis and Machine Intelligence. 9(6): 413～428

Chang S. K. and Jungert E., 1996. Symbolic Projection for Image Information Retrieval And Spatial Reasoning. London:

Academic Press

Chang S. K and Jungert E. and Li Y. 1989. Representation and retrival of symbolic pictures using generalized 2D-strings. In: SPIE Proc. Visual Communication and Image Processing. Philadelphia. PA

Chen J. et al., 1997. Improving 9-intersection model by replacing the complement with Voronoi region. In: proc. Pf Int. Workshop on Dynamic of Multi-dimensional GISs. Hong Kong

Chen X. Y., 1995. A workstation for three dimensional spatial data research. Proc. 4th International Symposium of LIES-MARS. Wuhan. 42~51

Christopher G. M. 1992. The meaning of neighbor. Lecture Notes in Computer Science 639. Pisa: Springer-Verlag. 220~235

Christopher G. M. and Yang W., 1996. Spatial date management tools based on the voronoi dynamic data model. User's Manual

Christopher G. M., 1997. The global GIS. In: Lee Y. C. and Li Z. L., ed. Proc. Int. Workshop on Dynamic and Multi-dimensional GIS. Hong Kong. 80~91

Chung M. C., Maria S. K. and Arctur D., 1995. An object-oriented approach for handing topology in VPF products. In: Proc. GIS/LIS'95. 163~174

Cui Z. Cohn A. G. and Randell A.A., 1993. Qualitative and topological relationships in spatial databases in design and implementation of large spatial database. In: Abel D. ed. Proc. 3rd Int. Sym. SSD'93. Pisa: Springer-Verlag. 396~415

Egenhofer M. J. and Herring J. A., 1990. Mathematical framework for the definition of topological relationships. In: Zurich K .B. and Kishimoto H. ed., Proc. 4th Int. Sym on Spatial Data Handling

Egenhofer M .J. and Franzosa R., 1991. Point-set topological spatial relations. Int. J. GIS. 5(2): 161~174

Egenhofer M. J. and Al-Taha K., 1992. Reasoning about gradual changes of topological relationships. In: Campari and Formentini U. ed. Theories and Methods of Spatio-Temporal Reasoning in Geographical Space. Lecture Notes in Computer Science 639. Pisa: Springer-Verlag. 196~219

Egenhofer M. J. and Mark D. M., 1995. Modelling conceptual neighborhoods of topological line-region relation. Int. J. GIS. 9(5): 555~565

Florence J. and Egenhofer M. J., 1996. Distribution of topological relations in geographic datasets. In: ASPRS/ACSM. Annual Convention and Expostion Technical Papers. 315~325

Frank A. U., 1992. Qualitative spatial reasoning about distance and directions in geographic space. J. Visual Languages and Computing. 3(2): 343~371

Frank A. U., 1996. Qualitative spatial reasoning: cardinal directions as an example. Int. J. GIS. 10(3): 269~290

Goyal R. K., 2000. Similarity Assessment for Cardinal Directions between Extended Spatial Objects. Ph.D thesis. The University of Maine

Kainz W., Egenhofer M. J. and Greasley I., 1993. Modelling spatial relations and operations with partially ordered sets. Int. J. GIS. 7(3): 215~229

Lee S. Y. and Hsu F. J., 1990. 2D C-string: a new spatial knowledge representation for image database systems. Pattern Recognition. (23): 1077~1087

Li Z. L., Zhao R. L., and Chen J. 2002. A Voronoi-based spatial algebra for spatial relations. Progress in National Science. 12(7): 528~536

Martien M., 1998. An Introduction to The Theory of Spatial Object Modelling for GIS. Talylor & Francis Ltd

Molenaar M. A., 1992. Topology for 3D vector maps. ITC Journal. (1): 25~33

Mukerjee A. and Joe G., 1990. A qualitative model for space. Proc. 8th National Conf. on Artificial Intelligence. Boston. MA. 721~727

Papadias D. et al., 1994. Topological relations in the world of minimum bounding rectangles: a study with R-tree. Technical Report. National Center for Geographic Information and Analysis. Santa Barbara

Papadias D. et al., 1995.. Topological relations in the world of minimum bounding rectangles: a study with R-tree. Proc. 1995 ACM SIGMOD Int. Conf. on Management of data. 92~103

Peuquet D. and Zhan C. X., 1987. An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane. Pattern Recognition. 20(1): 65～74

Pigot S. 1992. A topological model for a 3D spatial information system. Auto Carto. (10): 368～392

Pilouk M., Tempfli K. and Molenaar M., 1994. A tetrahedron-based 3D Vector Data Model for Geoinformation. In Advanced Geographic Data Modelling, (Ed., Molenaar M). Sylvia De Hoop: Geodesy Press. pp: 129～140

Pullar D and Egenhifer M., 1988. Towards formal definitions of topological relations among spatial objects. In: Marble D, ed. Proc 3rd Int. Sym. on Spatial Data Handling. 225～242

Randell D., Cui Z. and Cohn A., 1992. A spatial logic based on regions and connection. In: proc. 3rd Int. Conf. on Knowledge Representation and Reasoning. Boston. 165～176

Retz-Schmidt G., 1988. Various Views on Spatial Prepositions. AI Magazine, 9(2): 95～105

Safar M. and Shahabi C., 1999. 2D topological and direction relationship in the world of minimum bounding circles. Proc. 1999 Int. Sym. on Database Engineering And Applications

Shekhar S. and Liu X., 1998. Directions as a spatial object: a summary of results. Proc. 6th Int. Sym. on Advances in GIS. Washington. 69～75

Silvio L. 1997. Three-Dimensional Geometry And Topology. Princeton: Princeton University Press

Shi W. Z., Yang B. S. and Li Q. Q., 2003. An object-oriented data model for complex objects in three-dimensional geographical information systems. IJGIS (accepted)

Szmurlo M. and Gaio M., 1998.. Extended conceptual neighborhoods. In. Frirsh D. et al. ed. Proc. of ISPRS Commision IV Symposium. Stuttgart. 605～611

Worboys M. A., 1992. Geometric model for planar gepgraphical objects. Int. J. GIS. 6 (6): 353～372

Wu LX. Topological relations embodied generalized tri-prism (GTP) model for 3D geosciences modeling system (3D GMS) and Real-3D GIS. Computers & Geosciences. 2003(special issue, accepted)

Vieu L. A., 1993. Logical framework for reasoning about space. In: COSIT'93. Campari: Springer-Verlag. 25～35

Yang W. P. and Gold C. M. 1996. The VMO-tree: a dynamic spatial data model for maps. In: SDH'96. 260～270

Zhao R. L., Chen J. and Li Z. L., 1998. Voronoi-based generalized spatial adjacency. In: Li D. R. et al. ed. Proc. RS, GPS, GIS, Their Intergration and Applications. Wuhan Technical Uni. Surveying and Mapping Press. 605～614

第六章　时态 GIS 及时态关系

随着 GIS 研究与应用的不断深入，时态 GIS（Temporal GIS，简称 TGIS）的研究已成为 GIS 理论与技术研究的前沿领域和热点之一。TGIS 的组织核心是时空数据库（Spatio-Temporal Database，简称 STDB）（Langran G.，1992；张祖勋等，1995）。STDB 是时态数据库与空间数据库的结合，即对空间数据库赋予时态特征。通过将时间概念引入到 GIS 中，分析空间信息随时间的变换，描述系统在某时刻、时段或沿时间轴变化的过程，可以重现过去或预测未来状态，挖掘和发现系统沿时间变化的规律。

§6.1　时空过程与时空数据库

现实世界中一切事物，无论简单或复杂，均有一个从产生、到发展演变、最后消亡的过程。从 GIS 的角度，将现实世界的事物分为点、线、面、体以及简单对象、复杂对象等，统称为时空目标。TGIS 的任务就是对多时态及时空数据库（也称历史数据库），以动态的方式来描述时空目标的时空过程。

6.1.1　时空过程

时空过程实质是一系列沿时间轴的时空目标的变化过程，包括量变和质变。时空目标的描述包括几何、拓扑和属性三个基本方面，因此，时空变化包括沿时间轴的空间变化、拓扑变化和属性变化。图 6-1 是作者对时空过程的一个基本理解：

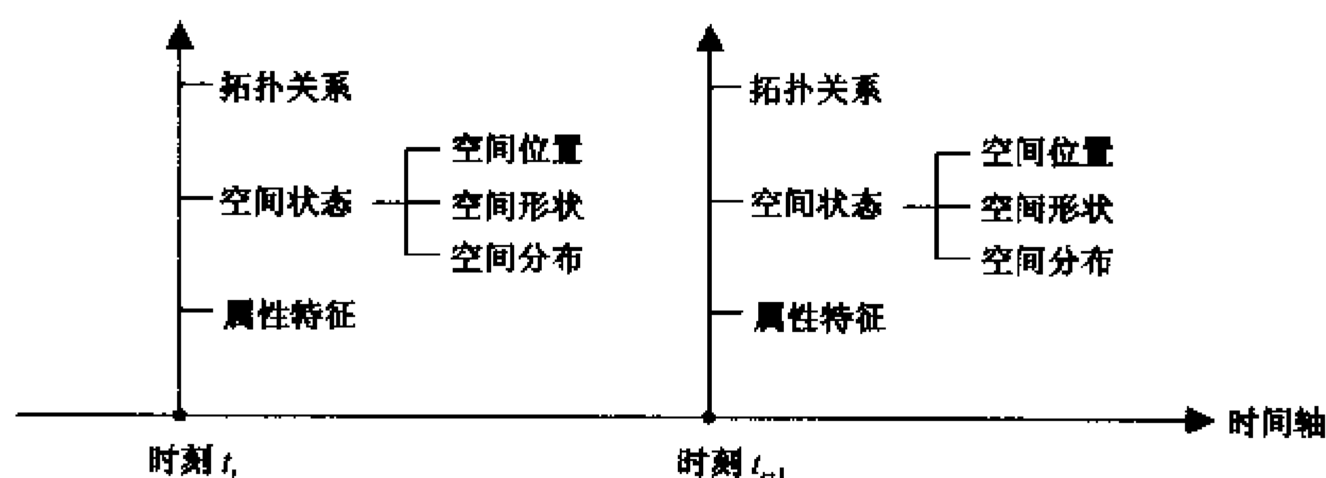


图 6-1　时空过程图解

事实上，后一时刻的拓扑关系、空间状态、属性特征均与前一状态的相应值有关系。在不同的时刻，空间对象的空间状态、拓扑关系和属性特征可能全部变化、两项变化或单项变化；空间状态的变化也可能是空间位置、空间形状、空间分布全部变化、或其中两项变化或单项变化。以地籍变更（包括土地分割、合并、所有权等）为例，一次地籍变更可能仅仅是所有权的变化，则属于属性特征变化；也可能既有土地的分割（或

合并)，也有所有权的变化，此时除属性特征变化之外，同等重要的是发生了空间状态的变化，包括空间位置和空间形状。此外，就地籍 GIS 而言，可能还牵涉到地块之间相邻关系的变化，即可能还发生了拓扑关系的变化。

尽管时空变化包括以上 3 项基本内容，但具体而言，目标的时空变化过程还是有其特定的主导变化方向。如：

(1) 以属性变化为主：基于行政边界的人口、资源、环境信息系统，空间目标的主体为各个不同级别的行政区，其几何状态是相对稳定不变的，而作为其属性的人口统计数据、资源勘测数据和环境数据等，则是动态变化的。

(2) 以空间变化为主：如交通运输信息系统和导航信息系统。导航系统的空间目标主体为车船飞机，其属性特征是相对稳定不变的，而其空间位置则是随时变化的。

(3) 空间变化与属性变化并重型：如矿山 GIS、地籍信息系统和气象信息系统。以气象信息系统为例，气象系统的空间目标的主体为风暴中心、高压云团等，监测的重点包括属性特征、空间位置、空间形状和空间分布。

空间过程的状态变化也分为连续变化和离散变化两种类型。如森林退化、草场沙化、湖泊退化、海岸盐碱化、海岸线变化、大气环境变化、城市化等属连续的量变的过程，可以在时间轴上进行内插；而地籍变更、道路改线等则属离散的质变，在时间轴上是不可内插的。

6.1.2 时空数据库

为了全面系统地描述空间目标的时空过程，记录空间目标从产生、演变到消亡全过程或阶段过程的历史数据是十分必要的。以土地纠纷为例，人们需要查找详细的历史记录作为法律裁判的依据。

记录时空目标的历史数据的数据库称为时空数据库（或历史数据库)。时空数据库是时态数据库（Temporal Database，简称 TDB）在 GIS 领域的扩展。TDB 作为数据库技术发展的一个重要分支，经历了 20 余年的发展历史，已经在管理、金融、人工智能等领域得到广泛的应用。

时空数据库有两个重要概念：其一为事件时间（event time)，指某一事件在现实时间中发生变化的时间；其二为系统时间（system time)，指数据库中跟踪记录事件发生的时间。所谓事件，“不但是时空目标状态开始或终结的标志，而且还表示状态变化的原因”(蒋婕等，2000)；林广发等（2002）进一步将驱动事件发生变化的直接原因狭义地定义为原子事件；而具有因果关联的多个原子事件组合为一个复合事件，它对应一个更宏观的对象的变化。

一般的非时态性 GIS 中，通常仅考虑事件时间。为了表达时空目标在数据库中的变化，有三种方式（Klopprogge M. R. T.，1981；Arain G. A.，1986；Gadia S. K.，1988)：①关系表重建型：当一个或若干个对象在一次事件中发生变化时，对这些对象所涉及到的关系表建立一个新的版本；②对象重建型：当一次事件变化后，仅对所涉及到并变化了的对象建立一个新的版本；③属性重建型：当一次事件变化后，仅对变化了的对象所涉及到的属性字段增加一个新值。

从数据冗余度考虑，显然，第三种方式最为理想。但是，由于该方式要求在一个属性字段中存放多个属性值，违背传统关系型数据库的第一范式（1NF）要求。因此，过去的许多研究均集中在第二种方式上。近年，随着面向对象（Object-Oriented，简称OO）技术的推广应用，为第三种方式的采用带来了新的希望。因为，OO模型打破了关系模型的范式要求，直接支持变长记录、多元组和子对象的嵌套。龚健雅（1997）提出为某些将会发生时态变化的字段进行特殊定义，允许它变长，即允许这种属性段在一个元组中嵌入多列多行的元组表，其中的行表达对象的多个版本，其中的列表达该对象的属性值及其时间说明。

时空数据库中与一个事件时间相对应的数据集称为一个事件版本，与一个系统时间相对应的数据集称为一个系统版本。龚健雅（1997）引入了一个新的概念，即对象版本，它是相对于对象自身的变化次数的数据集刻划。一次事件发生，整个数据库中各对象的系统版本时间是一致的，事件所涉及到的对象的事件版本也是一致的，而各对象的对象版本则不一致。因为，每个时空对象发生变化时时间和次数均是不统一的。由此，具有相同对象版本号的对象，所对应的变化时间可能并不一样；一个对象可能有多个不同版本的子对象组成。

§6.2　时空数据模型

时空数据模型是TGIS和STDB的基础。时空数据模型通常由数据结构、数据操作和完整性约束三部分组成（张祖勋等，1996）。TGIS不仅要处理属性数据和空间数据，还要处理时间维上的空属状态和空属变化，以及变化的过程与趋势；而且面向不同应用的TGIS，由于主要功能目的差异，需要选择不同的数据组织方式与时空数据模型。因此，围绕TGIS的空间、属性、时态三者之间的复杂关系和组织结构，人们研究提出了多种时空数据模型，概括起来，大致可以分为4类：①将时间作为属性的附加项；②将时间作为新的维数；③面向对象建模；④基于状态和变化的统一建模。除此之外，还提出了一些其他的时空数据模型，不一一讨论。

6.2.1　时间附加型

由于传统的关系型数据模型丰富的语义和较完善的理论基础，促使人们在最初阶段尝试扩充关系模型、关系代数和查询语言，试图将时间作为属性的附加项来模拟和支持时空数据的存储、表示和处理。关于属性数据的时态记录，数据库领域的研究多集中在第一范式（1NF）和非第一范式（N1NF）的讨论上。

1. 第一范式（1NF）模型

按第一范式（1NF）的要求，一个对象的历史过程需要用几个具有时间标记的元组来表达。在TGIS中，一个空间单元即使仅有属性特征的变化而无空间拓扑的变化，也必须用一个新元组来表示，从而在数据表中增加大量的数据冗余。

2. 非第一范式（N1NF）模型

非第一范式（N1NF）可以处理非表格化的复杂结构对象，元组可以采用不定长记录和嵌套方式。一般认为，N1NF是未来的发展方向（唐常杰，1999）。但N1NF的最大缺点是将同时发生的变化存储在树形结构的不同层次的结点上，给时态查询与时空分析带来了困难。陈军等（1995）针对这一问题，提出按变化发生的同时性将属性归并分组，得到经时间规范化处理后的N1NF关系表，从而达到以较小的冗余代价提高系统的效率。

3. 时空复合模型

时空复合（space-time composite）模型由Langran G.（1992）在前人研究的基础上加以发展的，其基础是时空单元。所谓时空单元，是指相同时空变化过程的最大单元。在存储上，每个时空单元被看成一个静态对象，其时空过程则作为属性用关系表来表达。时空复合模型将系统空间分割为若干时空单元，若时空单元发生分裂，则用新增的元组来反映新增的空间单元；时空过程每变化一次，即在关系表中新增一列时间段来表达。籍此，时空复合模型实现了用静态的属性表来表达动态的时空变化过程。该模型的缺点是数据库中的对象标识符的修改比较复杂，必须对标识符逐一进行回退修改，当涉及的关系链层次较多时，更显麻烦。

6.2.2 时间新维型

Berry B.（1964）最早提出使用时间维来描述空间对象的动态变化。在他提出的三维地理矩阵（geographic matrix）中，以位置、属性和时间分别作为矩阵的行、列和高。但曹志月等（2002）认为，由于高维对象构建算法的困难性等原因，还不能利用现有的GIS和DBMS来支持高维对象。

1. 立方体模型

该模型是以时间轴为发展变化方向，表达现实世界平面位置（二维图形）随时间的演变。给定一个时间位置，即可以从三维立方体中截取一个截面来得到该时刻的状态，如图6-2所示。将此模型由二维平面向三维空间扩展，同理可以表达三维空间状态随时间的变化。该模型的缺点是导致了数据量无限增长的趋势，而且随数据量的增大，对立方体的操作会变得越来越复杂，以致最后无法处理。

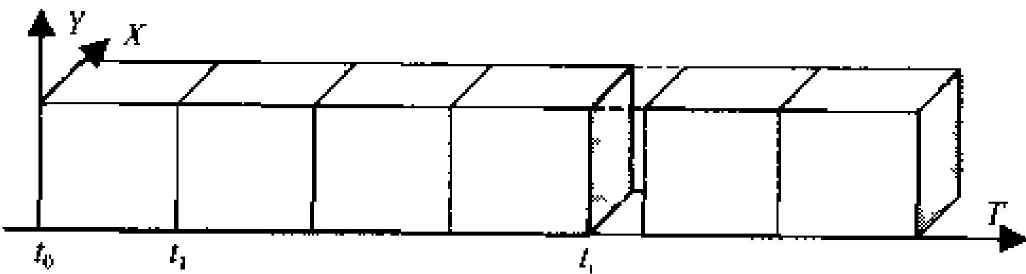


图6-2 时空立方体模型

2．系列快照模型

该模型是将一系列时间片断的快照（snapshot）存储起来，以此来记录系统空间状态的变化过程，如图6-3所示。需要时，可以根据指定的时间片段进行快照回放。由于每个快照均是对整个系统状态的不加取舍的全部记录，因此会产生大量的数据冗余，尤其当应用模型变化频繁时，系统效率会由于数据量激增而急剧下降。此外，该模型难以处理时空对象之间的时态关系。

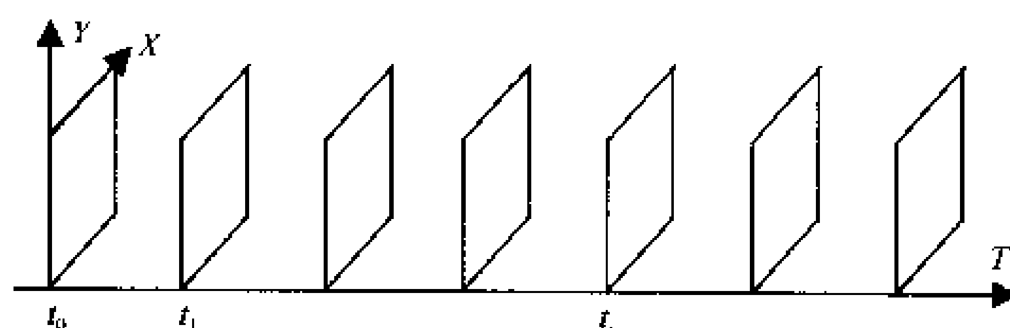


图6-3　时空系列快照模型

3．基态修正模型

基态修正（base state with amendments）模型是对系列快照模型的改进，即基于"差文件（deltafile）"（Kennedy-Smith G.M.，1986）的思想，选择某个时间的数据状态作为基态（base state）作完整记录，然后按预先设定的采样时间间隔，记录其他时刻相对于基态的变化量（amendments），以此来减少数据记录量。尽管基态修正模型可以较好地解决数据冗余的问题，但也给空间对象索引、空间关系表达和时空分析带来困难。

从操作速度来考虑，一般选择检索最频繁的状态作为基态。通常，系统的当前状态即最后一次更新的数据状态是用户最关心的，可以选择系统的当前状态作为基态。图6-4所示为4种典型的基态修正方式。

方式a把每个历史状态均作为对当前基态状态的修正。显然，每更新一次基态，就必须对所有的历史状态进行一次更新以求得新的差文件。当历史久远或历史状态众多时，这是不现实的。

方式b（Langran G.，1988）的特点是每个早期的历史状态均是对其紧后的一次历史状态的修正，可以从当前基态开始，通过逐步回溯序列差文件的方式来重构每一个历史状态。显然，这一方式虽然避免了方式a的缺陷，但也给历史状态的操作带来困难。

方式c（张祖勋等，1996）是在方式b的基础上，选择一些特定的重要历史状态直接与当前基态进行方式a的修正，形成另一组更宽时间跨度的差文件序列。

而方式d（张祖勋等，1996）则是在方式b的基础上，叠加一组具有更宽时间跨度的方式b的差文件序列。这两种方法虽然可以在一定程度上弥补方式a和方式b的不足，但当历史久远时，依然存在方式a和方式b中的问题。

刘仁义等（2000）提出以现时库、过程库和历史库（三者之间的逻辑关系如图6-5所示）来扩展基态修正模型，考虑到现时库中每个元组均处于激活状态，是数据库的当前操作对象，建议以现时库作为基态库，并以此为基础设计了一套商业化的土地产权产籍管理系统。

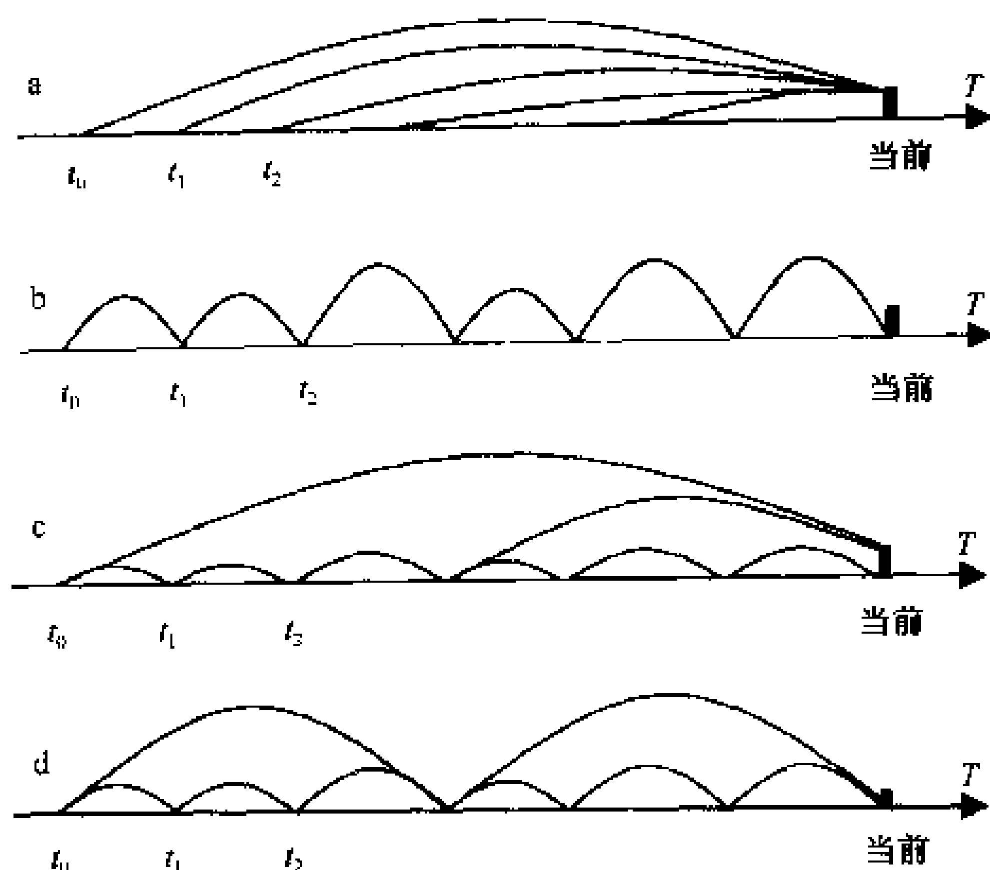


图 6-4　基态修正的 4 种典型方式
（据曹志月等，2002）

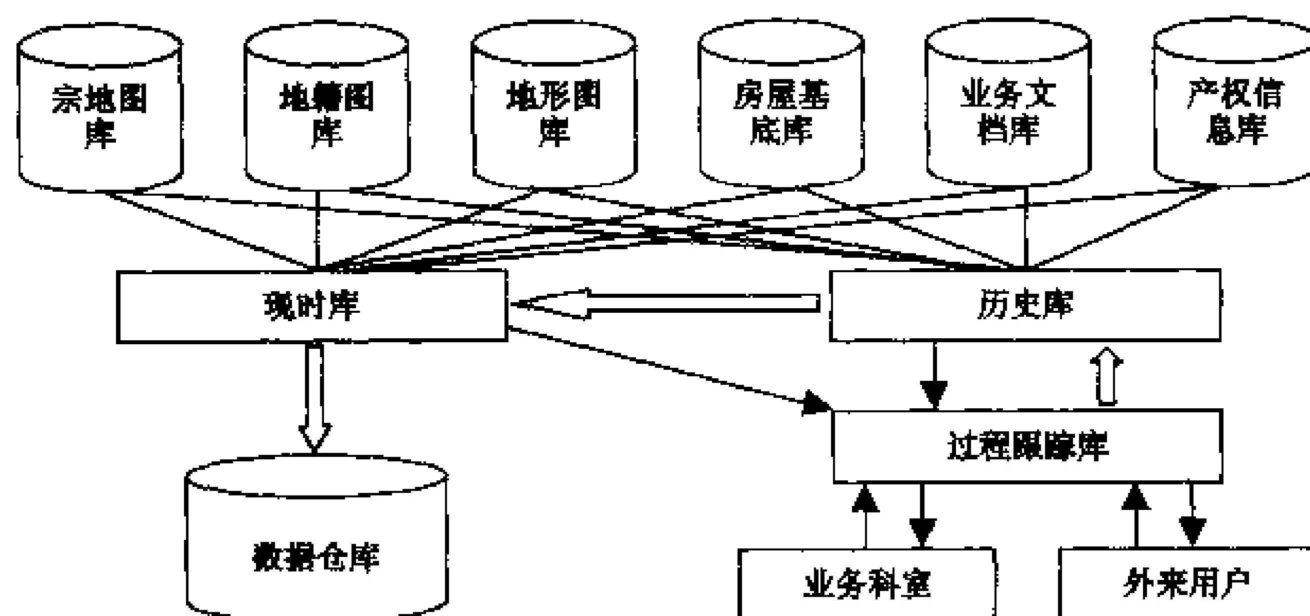


图 6-5　现时库、过程库、历史库的逻辑关系
（刘仁义等，2000）

作者认为，照顾当前操作速度固然重要，但数据库维护也不可忽视。如果以系统的当前状态或现时库作为基态，显然基态是一个变量；随着基态的动态变化，前述的基态修正方式 a、方式 c、方式 d 均牵涉对有关历史状态的修正，即必须随时对历史库进行维护和更新，这几乎是不可容忍的。鉴此，选择过去的某个时刻（尽可能靠近当前）的数据相对完整、数据质量好的状态作为基态是一值得考虑的思路。

曹志月等（2002）等提出动态多级索引的方式来进行基态修正。如图 6-6 所示，该方法的思路是：在整个历史状态中动态地建立多个基态，用户根据具体需要确定基态距（即各基态之间的差文件数）阈值；系统随着新的历史状态的形成，即新的差文件的插入和修正，根据对基态距的判断而动态地创立多级基态。显然，这一方式可以提高数据

的存取效率，并在一定程度上摆脱历史久远情况的影响。该方式的关键是确定合理的基态距阈值，阈值太小会浪费存储空间，阈值太大则会占用过多的时间进行历史状态回溯与重构。杜道生等（1997）在研究碎分拓扑弧段的时间标记方法问题时，曾提出设立多个基态并由用户依据应用要求给出分裂容限值（split-tolerance）的措施来控制弧段的分裂过程。

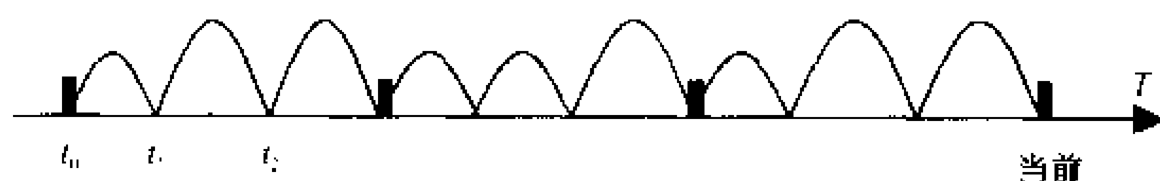


图 6-6 动态多级索引的基态修正方式
（曹志月等，2002）

6.2.3 面向对象型

尽管 OO 技术本身在建模概念、理论基础和实现技术等方面还不够成熟，但由于它能以更自然的方式对复杂时空对象进行模型化，可以打破关系模型的范式要求，直接支持对象嵌套和变长记录，使得 OO 技术在 TGIS 研究领域备受关注。Worboys M. F.（1992）、Peuqust D. J. et. al（1995）、Shi W. Z.（1995）、龚健雅（1997）、林广发等（2002）、曹志月等（2002）分别围绕 OO 型时空数据模型进行了研究。

曹志月等（2002）提出按以下对象结构描述时空对象的时间、空间和属性 3 个基本成分：

$$<OBJ:\{O\text{-}ID, Attr(t), Spatial(t), Temporal(Tv,\ Td), Actions\}> \qquad (6\text{-}1)$$

其中：

OBJ——时空对象，可以是点类、线类、面类、体类、简单对象类或复杂对象类；

O-ID——对象的惟一表示符；

Attr（t）——对象的随时间 t 变化的非空间属性描述；

Spatial（t）——对象的随时间 t 变化的空间特性描述；

Temporal（Tv，Td）——对象的时态性描述，反映对象的产生、演变和消亡的生命历程，其中有效时间 Tv 与数据库 Td 时间是正交的；

Actions——对象的行为操作描述，其中定义对象的时间、空间及属性的各种运算操作，实现同类对象或不同类对象之间的相互联系，通过 OO 的多态性和继承性，以自然简洁的方式实现运算符和拓扑集合关系的重载。

曹志月等（2002）所定义的系统时空对象系统中，每个对象对应一个惟一的有效时间链表和惟一的数据库时间链表。其中，有效时间链表描述该对象在现实世界中的实际变化历史，链表按对象的 Tv 顺序排列，采用基于基态的动态多级索引方式存储。数据库时间链表描述该对象在数据库中被表示、更新与处理的历史过程，链表按对象的 Td 顺序排列，Td 顺序是由系统的操作处理时间确定的。

林广发等（2002）则提出以事件为核心的时空数据模型，并以房地产信息系统为例进行了设计实现。该模型的要点是：①为每一类对象定义一个“内核”来表示其本质，其标识与对象标识一致；②归纳提取驱动各类对象发生变化的原子事件的集合，定义这

些事件将导致哪些对象的哪些属性同时发生变化，以及如何变化；③将对象的属性（含空间图形单元）按变化的同时性（即“时态内聚性”）分组存储，每组为一个数据管理的基本单元；④将事件归类编码，并将时态信息用“时间＋事件编码”字符串标记在各基本单元上；⑤定义事件之间的连锁关系，以事件为核心建立数据完整性控制表，并据此驱动数据的更新过程。

6.2.4 基于状态和变化的统一模型

郑扣根等（2001）认为，面向对象的时空数据模型存在两方面的缺陷：①在构造数据时，没有按语义来进行结构化的数据组织，对象的时空属性与其他属性没有区别；②缺乏时空变化的语义，不能表达空间拓扑关系的变化。而实际上，在对真实世界进行建模时，空间、时间和对象的独立性均包含了丰富的语义。鉴此，郑扣根等（2001）对时间和空间语义进行了抽象和形式化，定义了时空数据库中的状态、变化和事件的概念，进而提出了一个基于状态和变化的统一时空数据模型 SCUDM（State and Change based Unified Data Model）。

SCUDM 是在时空复合概念和时空单元（Langran G.，1992）的基础上发展起来的，并对时空元素、空间实体、时空对象、时空数据库状态和时空数据库变化进行了以下定义（郑扣根等，2001）：

定义 1：时空元素是时空立方体范围内的最大的共同时空单元。

定义 2：空间实体是在时空立方体范围内一定时间段上不变的空间状态。

定义 3：时空对象是时空数据库中的语义对象。它由语义上的特征标识自己，而与特定的时间和空间状态无关，即时空对象与具体的空间和时间相分离。

定义 4：时空数据库的状态是一个三元组：

$$\text{state} := (o, s, t), o \in \text{OBJ}, s \in \text{Space}, t \in \text{Time} \tag{6-2}$$

其中 OBJ、Space、Time 分别是时空数据库的对象域、空间域和时间域。

定义 5：时空数据库的变化是关于数据库状态的变化，是从一个数据库状态到另一个数据库状态的函数 f^{OST}：state→state，f 的上标 OST 表示状态可以在对象域、空间域和时间域之间变化。变化的过程是一系列的变化 F^{OST}：f_1^{OST}，f_2^{OST}，…，f_m^{OST}。

由于状态的变化可能会很复杂，即状态三元组中的对象变量、空间变量和时间变量均可能在各自的域内发生变化。为了方便描述这种复杂变化，郑扣根等（2001）把 f^{OST}和 F^{OST}从 3 个域上的变化降到两个域上的变化，即将状态的变化分别沿对象域、空间域和时间域进行投影。从而，将时空数据库的变化分解为以下 3 个二维变化：

$$\begin{cases} \text{对象域上的变化}: f^{\text{ST}}: \text{state}\mid_O \rightarrow \text{State}\mid_O, F^{\text{ST}}: f_1^{\text{ST}}, f_2^{\text{ST}}, \cdots, f_m^{\text{ST}} \\ \text{空间域上的变化}: f^{\text{OT}}: \text{state}\mid_S \rightarrow \text{State}\mid_S, F^{\text{OT}}: f_1^{\text{OT}}, f_2^{\text{OT}}, \cdots, f_m^{\text{OT}} \\ \text{时间域上的变化}: f^{\text{OS}}: \text{state}\mid_T \rightarrow \text{State}\mid_T, F^{\text{OS}}: f_1^{\text{OS}}, f_2^{\text{OS}}, \cdots, f_m^{\text{OS}} \end{cases} \tag{6-3}$$

SCUDM 采用以下 3 种模式（Domain schema，DS）来表达对象域、空间域和时间域上的状态变化：

$$
\begin{cases}
\text{对象域模式:Object-DS:} =< o, < s_1, t_1 >, \cdots, < s_n, t_n >>; n \geqslant 1, n \in N \\
\text{空间域模式:Space-DS:} =< s, < o_1, t_1 >, \cdots, < o_n, t_n >>; n \geqslant 1, n \in N \\
\text{时间域模式:Time-DS:} =< t, < o_1, s_1 >, \cdots, < o_n, s_n >>; n \geqslant 1, n \in N
\end{cases} \tag{6-4}
$$

如图 6-7 所示，SCUDM 构成一个由上至下由 Object-DS-layer、Time-DS-layer 和 Space-DS-layer 组成的 3 层数据模型。时空对象位于模型的最上层即 Object-DS-layer，其空间属性指向多个时空实体，该层采用时空对象表（st _ object _ table）描述时空对象的空间变化，表中的每一条记录收集一个时空对象在其生命周期（lifespan）内的所有历史版本。Time-DS-layer 采用时空实体表（st _ entity _ table）描述时间上的空间-对象变化，在同一时间点上的时空实体构成此时间的静止图层；时空实体的几何描述并不在静止图层内表达，而是指向多个时空元素，因而时空实体只是一个代理。Space-DS-layer 采用时空多边形表（st _ polygon _ table）来描述空间几何形状上的变化，表中的每一条记录是关于一个时空元素的时间序列信息，描述该时空元素所表达的空间区域在不同的时间为不同的时空实体所占用的变化情况；同时，通过时空实体这一代理来与时空对象发生联系。由时空元素构成的复合图层是一个单一的图层，它把多个时间点上的静态图层复合在一起，并承担空间几何描述和拓扑描述的任务。

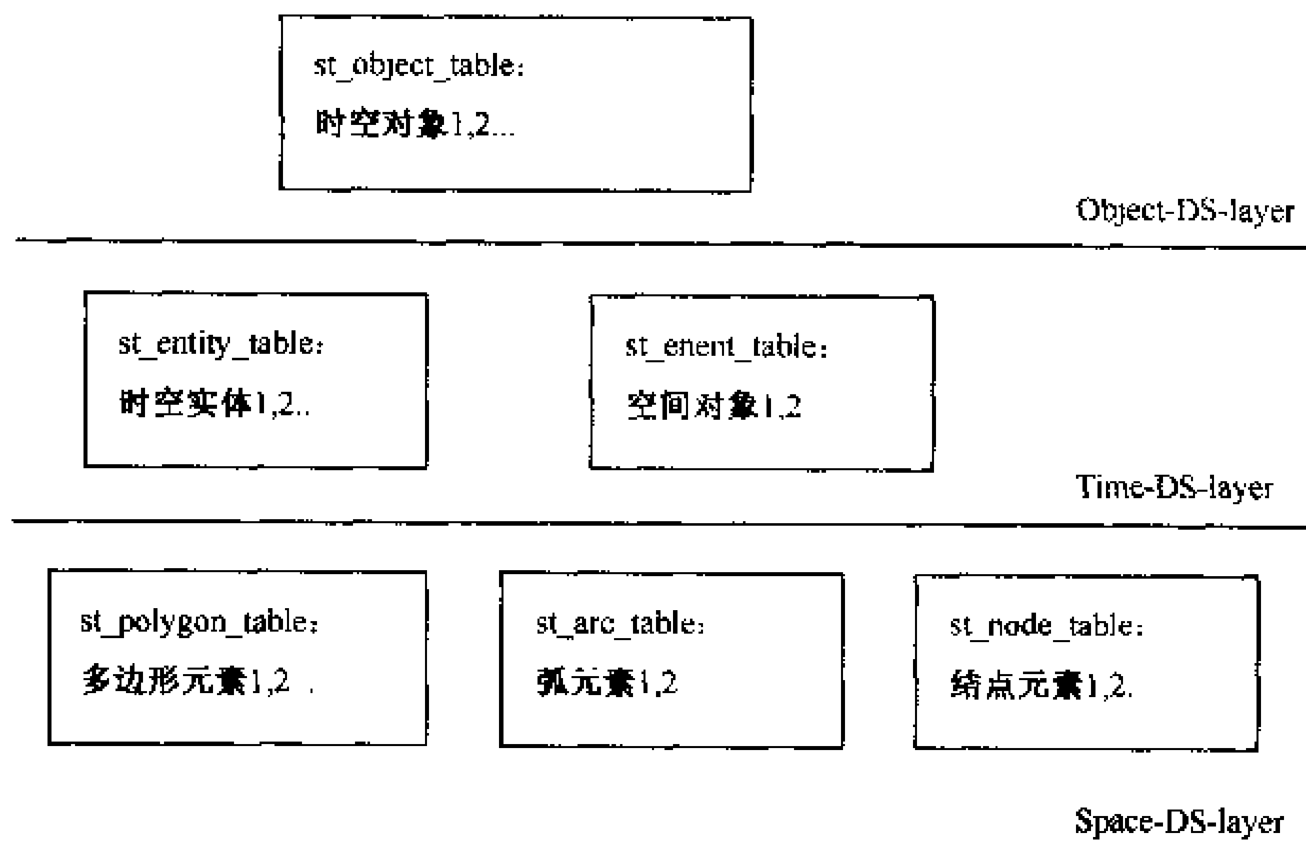


图 6-7 SCUDM 数据模式
（据郑扣根等，2001）

6.2.5 其他时空数据模型

杜道生等（1997）鉴于现实世界中存在许多同步变化的细小单元，而传统的数据模型难以适应这种细节的表达，并且在数据冗余和局部信息提取方面存在诸多问题，研究提出了一种基于属性同步数据项组合空间碎分拓扑弧段时间标志的时态地理数据模型。该模型的基本思想是：将研究范围内同步变化的若干最小单位（属性数据项和空间坐标

点）归为一组，形成若干同步数据项组和同步坐标点组，并分别附上一个时间标记。当实体内全部单元值变化同步时，该时间标记演化为实体时间标记；当实体内各数据单元变化节奏互不相同时，该时间标记演化为最小单元时间标记。基于这样的处理，可以达到既体现地理状态的局部变化，又最大限度地避免不变状态的数据冗余存贮。

张显峰等（2000）则鉴于元胞自动机（Cellular Automation，简称 CA，由离散的元胞（cells）、有限的状态（states）、邻域（neighbor）和规则（rules）4 个基本元素组成）是一种基于元胞或单元的"自上而下"的动态建模框架，并具有模拟复杂系统时空演化过程的能力，认为可以与基于栅格结构的 GIS 结合起来，增强 TGIS 的时空动态建模功能。进而以土地利用为例，对 CA 模型进行了元胞空间扩展、元胞状态扩展、元胞转换规则扩展和时间概念扩展，并基于扩展的 CA 模型研究建立了基于 GIS 与 CA 集成的土地持续利用时空模拟，设计了时空模拟引擎，如图 6-8 所示。

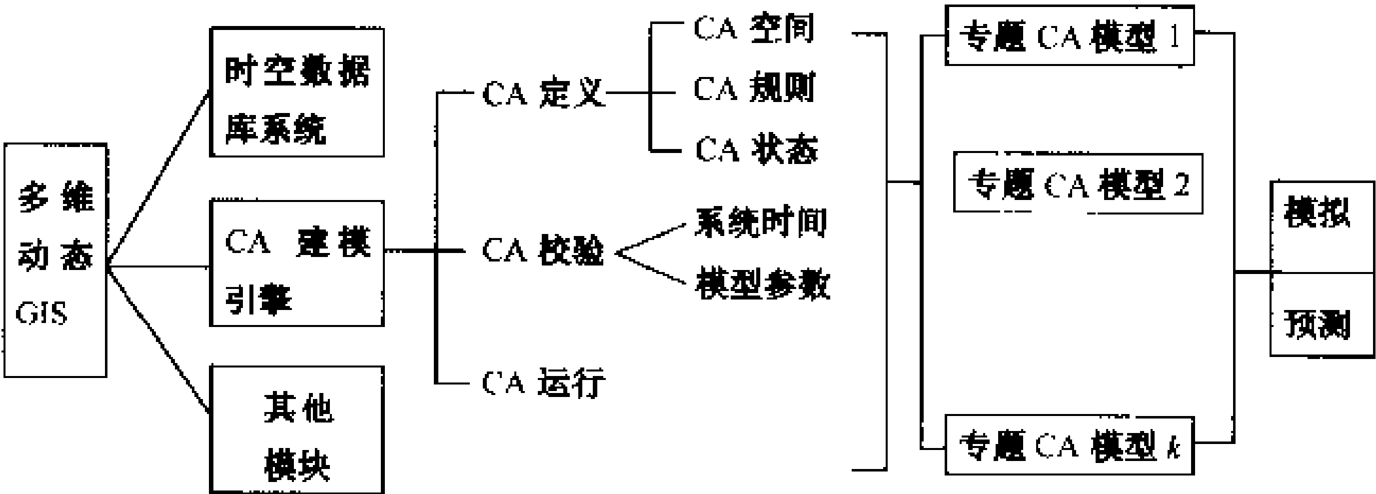


图 6-8 集成 CA 与 GIS 生成时空模拟引擎
（据张显峰等，2001）

陆锋等（2001）分析，传统的基于专题的地理分层的空间数据表达思想和图层内基于矢量或栅格的建模单元表达方式，割裂了各层次要素之间可能存在的语义与空间关系，难以表达复杂地理实体和描述地理过程。指出：为了客观真实地对现实世界进行概括，必须从几何分布、专题要素与时态变化 3 个方面来描述地理特征。进而提出发展基于地理特征（feature）的时空数据模型的必要性，并讨论了基于特征的概念模型、相互关系与关键技术。

George C. 等（2002）将现代时空统计学中的 Bayesian Maximum Entropy（BME）技术引入现今 TGIS 技术中，发展和改良了 TGIS 的时空分析与建模能力，为自然、生态和社会经济决策与规划提供了强有力的工具。

§6.3 时态拓扑关系

时态拓扑关系（有时也称时空拓扑关系）是时态 GIS 的一项重要内容。刻划时空实体之间的联系，可以帮助人们以相关关系为核心，从本质、宏观上抽象把握客观世界的相互关联规律。

6.3.1 时态拓扑关系的定义

迄今，国际上关于时空拓扑关系的定义还没有一个完整的、普遍接受的定义。众多定义中，归纳起来，主要有3类不同的出发点：

其一：从数学模拟的角度出发（Armstrong M.P.，1988；Altaha K. K. et al.，1993），类似于空间拓扑关系的定义，该定义并未揭示时态拓扑的内涵及其与空间拓扑的关系；

其二：从时间顺序出发，或针对事件（Langran G.，1992；Peuquet D. J.，et al.，1995；黄明智等，1997；郑扣根等，2001），或针对状态（H. Raafat et al.，1994），实际混淆了时间方向与时态拓扑的关系；

其三：从点集理论出发（舒红等，1997），借助点集理论描述时空拓扑关系，并给出相应的谓词描述。

以从时间顺序出发针对事件的定义为例，按照时空实体的属性、空间几何和空间拓扑三个基本要素，可以将时空事件区分为属性事件（属性发生改变的事件，简称Ae）、空间几何事件（空间几何发生改变而拓扑关系维持不变的事件，简称SGe）和空间拓扑事件（拓扑关系发生改变的事件，简称STe）。舒红等（1997）认为，时态拓扑关系刻划属性事件、空间几何事件和空间拓扑事件之间的拓扑关系，而时空拓扑关系则主要指空间拓扑事件之间的拓扑关系，如图6-9所示。

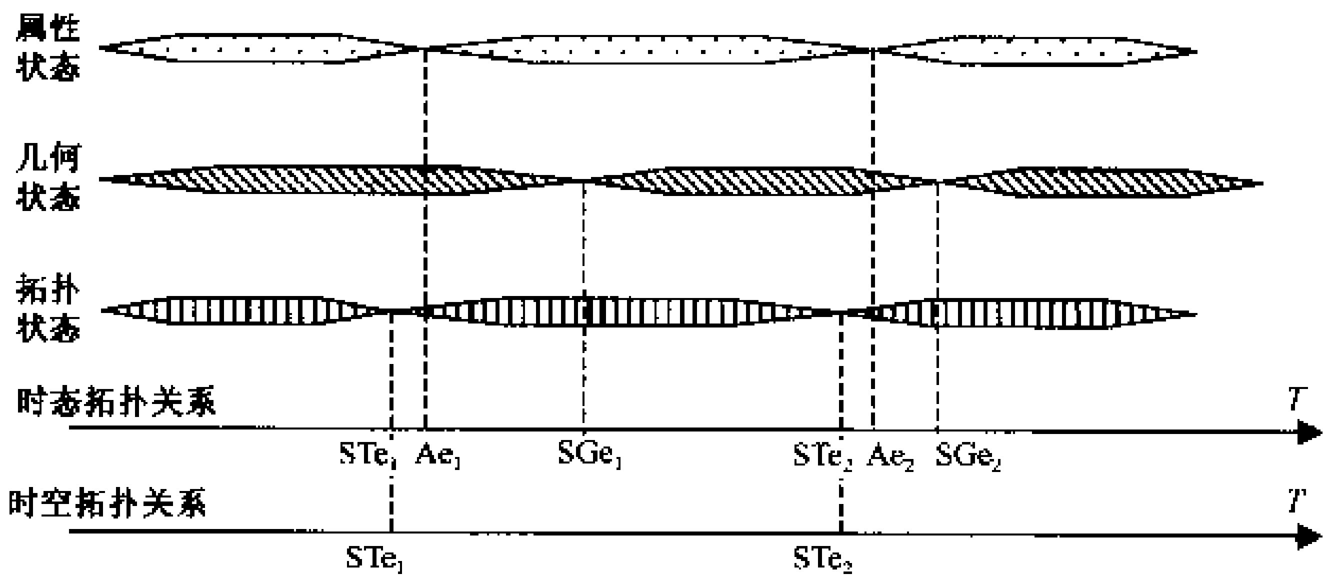


图6-9 基于事件的时态拓扑关系图形表示
（据舒红等，1997）

图6-9表示的是同一时空实体的状态变化引起的该实体的时态拓扑关系的改变。此外，还包括引起不同地理实体状态变换的两个事件之间的时态拓扑关系，如图6-10所示。一个地理实体的属性状态变化可以进一步区分为产生（Birth）、进化（Evolution）和消亡（Death），所对应的属性事件分别简称为ABe、AEe和ADe。

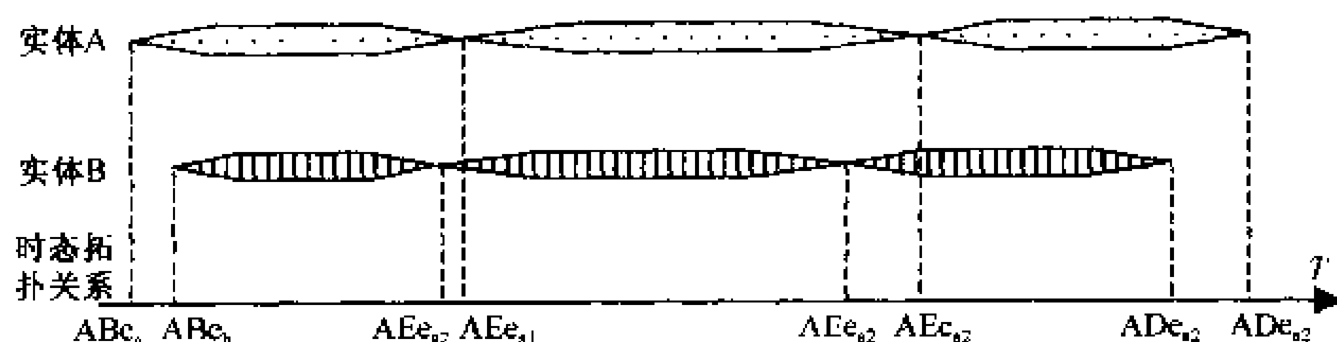


图 6-10　基于事件的时态拓扑关系图形表示
（据舒红等，1997）

6.3.2　时态拓扑关系的描述

时间可以区分为绝对时间定位值和相对时间关系。其中，绝对时间是标识事件的起始和终止的时刻位置，如事件 I 的起始时刻 T_{snode}可以描述为 Begin（I）＝T_{snode}，而其终止时刻 T_{enode}则可以描述为 End（I）＝T_{enode}。相对时间关系主要指时间方向、时间距离和时空事件的时间拓扑关系。时间方向关系有“先”（before）、“后”（after）和“同时”（equal），如 Before（I，J）表示事件 I 先于事件 J 出现（或称 $J \leqslant I$），而 After（I，J）则表示事件 I 后于事件 J 出现（或称 $I \leqslant J$）。时间距离关系有“早于”（early）、“晚于”（later）；时间拓扑关系有“相离”（disjoint）、“相遇”（meet）、“叠加”（overlap）、“覆盖”（cover）、“被覆盖”（covered by）、“相等”（equal）、“内部”（inside）、“包含”（contain）等。

过去，人们常将时态拓扑关系和时间方向关系结合起来研究，Allen J. F.（1983）所提出的 12 种时间区间关系（当时未考虑 Equal）即是其中的典型，如表 6-1 所示。

表 6-1　时间区间关系

（据 Allen J. F.，1983）

图示	I J	I J	I J	I J	I J
区间关系	Before（I，J）	Meet（I，J）	Overlap（I，J）	Start（I，J）	Startedby（I，J）
图示	J I	J I	J I	I J	I J
区间关系	After（I，J）	MetBy（I，J）	OverlappedBy（I，J）	Finish（I，J）	Finishedby（I，J）
图示	I J	I J	I J		
区间关系	Equal（I，J）	During（I，J）	Contain（I，J）		

舒红等(1997)将时态目标所处的时间背景称为时间全集 T，并将时间全集 T 模拟成带“$\leqslant$”全序关系的一维实数欧氏空间 IR^1，其维数恒为1，即 $\mathrm{Dim}(IR^1)=1$。通过证明四元组框架(4I)的拓扑关系描述能力等价于九元组加维数扩展框架(9I＋DEM)的拓扑描述能

力，进一步推论："时间全集里时态目标间的时态拓扑关系可以采用基于点集理论的四元组框架41完备、惟一地描述，且存在与之等级的最小谓词描述集{disjoint，touch，cross，overlap，in}"。

从数学理论可知，四元组的每一项可以为空∅（Dim（∅）＝－1）或非空¬∅（Dim（¬∅）≡1）。因此，四元组共有$2^4=16$种可能的取值情况，即有16种可能的时态拓扑关系情形。舒红等（1997）根据时间全集和时态目标的实际情况，剔除了8种不能成立的情形，得到如表6-2所示的8种时态拓扑关系情形。

表6-2 两时态目标之间的时态拓扑关系

（舒红等，1997）

图示	I J	I J	I J	I J	I J	I J	I J	I J
四元组	$\begin{bmatrix}\varnothing & \varnothing\\ \varnothing & \varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \varnothing\\ \varnothing & \varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \neg\varnothing\\ \neg\varnothing & \neg\varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \varnothing\\ \neg\varnothing & \neg\varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \neg\varnothing\\ \varnothing & \neg\varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \varnothing\\ \varnothing & \neg\varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \neg\varnothing\\ \varnothing & \varnothing\end{bmatrix}$	$\begin{bmatrix}\neg\varnothing & \varnothing\\ \neg\varnothing & \varnothing\end{bmatrix}$
谓词	$T_{disjoint}$	T_{meet}	$T_{overlap}$	T_{cover}	$T_{coveredBy}$	T_{equal}	T_{inside}	$T_{contain}$
语义	I、J间隔出现	I、J相遇出现	I、J部分同时出现	I变化期间J出现	J变化期间I出现	完全同时出现	I在J出现过程中出现	J在I出现过程中出现

对比表6-1和表6-2可见，两者之间可以建立如表6-3所示的对应关系。

表6-3 两时态目标之间的时态拓扑关系与时间区间关系的对比

舒红等，1997	$T_{disjoint}$	T_{meet}	$T_{overlap}$	T_{cover}
Allen J.F.，1992	Before（I，J） After（I，J）	Meet（I，J） MetBy（I，J）	Overlap（I，J） OverlappedBy（I，J）	StartedBy（I，J） FinishedBy（I，J）
舒红等，1997	$T_{coveredby}$	T_{inside}	T_{equal}	$T_{contain}$
Allen J.F.，1992	Starte（I，J） Finish（I，J）	During（I，J）	Equal（I，J）	Include（I，J）

参考文献

曹志月，刘　岳．2002．一种面向对象的时空数据模型．测绘学报．31(1)：87～92

陈　军，陈尚超，唐治锋．1995．用非第一范式关系表达GIS时态属性数据．武汉测绘科技大学学报．20(1)：12～17

杜道生，舒　红．1997．基于同步数据项组合碎分拓扑弧段时间标记得时态地理数据模型．武汉测绘科技大学学报．22(2)：98～101

龚健雅．1997．GIS中面向对象时空数据模型．测绘学报．26(4)：289～298

黄明智，张祖勋．1997．时空数据模型的N1NF关系基础．测绘学报．26(1)：1～6

蒋　捷，陈　军．2000．基于事件的土地划拨系统的时空数据库设计．测绘学报．29(1)：65～70

林广发等．2002．以事件为核心的面向对象时空数据模型．测绘学报．31(1)：71～76

刘仁义，刘　南，苏中国．2000．时空数据库基态修正模型的扩展．浙江大学学报(理学版)．27(2)：196～200

陆　锋等．2001．基于特征的时空数据模型：研究进展与问题探讨．中国图象图形学报．6A(9)：830～835

陆汝钤，金　芝，陈　刚．2000．面向本体的需求分析．软件学报．11(8)：1009～1017

舒　红等. 1997. 时空拓扑关系定义及时态拓扑关系描述. 测绘学报. 26(4)：299～306

舒　红,陈　军,史文中. 1998. 时空数据模型研究综述. 计算机科学. 25(6)：70～74

唐常杰. 1999. 时态数据库的成果、缺陷与未来. 计算机科学. 26(3)：63～65

张师超. 1992. 时态数据库述评. 计算机科学. 19(3)：37～43

张师超. 1993. 时态关系代数与元组演算的等价性. 计算机学报. 16(12)：936～939

张显峰,崔伟宏. 2000. 集成GIS和CA模型的时空建模方法研究. 中国图象图形学报. 5A(12)：1012～1018

张显峰,崔伟宏. 2001. 集成GIS和细胞自动机模型进行地理时空过程模拟与预测的新方法. 测绘学报. 30(2)：148～155

张祖勋,黄明智. 1995. 时态GIS的概念、功能和应用. 测绘通报. (2)：12～14

张祖勋,黄明智. 1996. 时态GIS数据结构的研讨. 测绘通报. (1)：19～22

郑扣根,谭石禹,潘云鹤. 2001. 基于状态和变化的统一时空数据模型. 软件学报. 12(9)：1360～1365

Allen J. F., 1983. Maintaining knowledge about temporal intervals. Communications of the ACM. 26(11)

Al-Taha K.K. et al., 1993. Bibliography on spatiotemporal database. ACM SIGMOD Record. 22(1): 59～67

Arian G. A. 1986. A temporary oriented data model. ACM transactions Database Systems. (11): 499～527

Armstrong M. P. 1988. Temporality in spatial databases. In: Proc. Pf GIS/LIS'88. Falls Church, VA: ACSM, Volume 2: 880～889

Berry B. 1964. Approaches to regional analysis: a synthesis. Annals of the Association of American geographers. 54: 2～11

Gadia S. K. 1988. A homogenous relational model and query languages for temporal databases. ACM transactions Database Systems. (13): 418～448

Gadia S. K., Yeung C. 1991. Inadequacy of interval time stamps in temporal database. Information Science. 54(1～2): 1～22

George C., Patrick B., Marc L.S., 2002. Temporal GIS: Advanced Functions for Field-Based Applications. Germany: Springer-Verlag

Kennedy-Smith G. M. 1986. Data quality: A management philosophy. Proc. Of Auto Carto. London. 381～390

Klopprogge M. R. T. 1981. An approach to introduce the time dimension in the entity-relationship model. In: GIS. Int. J GIS. 9(1): 7～24

Langran G. 1988. temporal GIS design trade-offs. Proc. Of GIS/LIS'88. San Antonio. 890～899

Langran G. 1992. Time in GIS. Bristol, PA: Toylor & Francis

Matthew Y C P., Shi W. Z. 1998. Process-based spatial-temporal model: handing spatial and temperal pattern in a unique model. Proc. of Geoinfomatics'98 Conference.

Peuguet D. J., Niu D. 1995. An event based spatio-temporal data model (FSTDM) for temporal analysis of geographical data. Int. J. GIS. 9(1): 7～24

Raafat, H. et al. 1994. Relational spatial topologies for historical geographical information. International Journal of Geographic Information Systems, 8(1): 95～103

Shi W.Z. 1995. Object-orinted Approch for spatial, temporal and attribute data modeling. Proc. Of GIS/LIS' 95. Nashiville

Worboys M. F. 1992. Object oriented models of spatio-temporal information. Proc. Of GIS/LIS'92. Atlanta GA: ACSM. 825～834

第七章　地学信息分类与空间编码

广义上讲，地学信息是指通过测量、遥感、地理、地质、地球物理、地球化学等多学科手段获得的关于地球表层及其内部的各种信息。这些信息是进行地球科学与区域性科学研究的基本素材，特别是自然资源、环境评估、灾害预测、国土规划、工程决策和可持续发展研究的基本依据。建立地学信息的分类与编码体系既是地学信息标准化的一项重要内容，也是GIS系统建设的前期基础工作。从GIS的角度，地学信息的分类和编码包括两项主要内容，即地学属性信息的分类与编码和地学空间实体的地址编码。其中，属性信息的分类编码主要是为了数据格式的统一和数据共享；而空间实体的位置编码主要是为了空间查询和位置检索，这在第二章、第三章中种已有相应介绍。关于地学信息分类与编码，国际上多有研究，ISO-TC211有专门的工作组从事这方面的工作，我国也开始制定一些相应标准。本章将重点结合目前非3D GIS理论与技术现状来讨论地学属性信息的分类与编码问题，并介绍作者面向下一代真3D GIS的地址编码的研究成果。

§7.1　分类编码原则与标准

7.1.1　分类编码原则

地学信息分类编码的一般原则是：科学性、系统性和实用性。

1．科学性原则

主要指分类编码依据的合理性、分类体系的完备性和学科综合性。

首先，“要选择事物或概念（分类对象）最稳定的本质属性或特征作为分类的基础和依据”（GB7027-86）。比如，地学各个分支学科经过长期发展均形成了各自的相对稳定的研究对象、研究方法和信息获取手段，因而可以将地学信息的“信息来源”这一属性视为一种稳定属性进行考虑。

其次，分类体系应能全面、完整地反映地学信息的客观状况。由于地学属性值既有明确定量的数据（如地面坡度15°），非明确定量或模糊描述的信息（如约15°，近水平），也有定性描述的信息（如厚层、直立等），因而在对属性进行分类时，各种数字和文字描述数据都能找到它们的对应位置。

再次，随着学科交叉与融合，多源地学信息的集成是一个不可逆转的趋势。以矿产资源评价为例，地质、物探、化探、遥感信息的综合日益发挥重要作用。因此，在建立地学信息分类编码体系时，应对研究目标所涉及的各有关学科的信息从整体上进行综合考虑，纳入同一系统，以便现在和将来的综合应用。以美国1993年出台的“1∶5万～1∶10万地质图图形分类编码标准”（征求意见稿）为例，采用8位码，其中首位码为数

据大类码，整个地学数据（Earth Science Data）是其中的一个大类。

2．系统性原则

主要指体系内各部分之间的有机联系与和谐统一的关系，强调系统的整体性、共性的归纳和结构的简明性。一是要将多源地学信息纳入统一的分类体系，使体系丰满完整；二是要通过归纳、抽象手段，从属性类别中寻找规律、共性及其联系，以便建立类别之间的系统联系；三是要在“完备性”前提下和共性研究的基础上，选择合适的分类准则，去掉冗余信息，使分类体系结构尽量简化。

3．实用性原则

主要指所涉及的地学信息分类编码体系不仅要方便地学数据库的建设，还要有利于数据库的规范化管理。这取决于分类的科学性和系统性，以此来保障一般数据库系统的关于数据独立性的要求，从而便利于数据的操作和维护。

7.1.2 分类编码标准

在地学信息分类编码方面，世界各国均制定了相应的标准，以此来规范数据建设和促进数据共享。

以美国地质调查局（USGS）于20世纪80年代末90年代初制定的《数字线划图形标准》为例，该标准采用7位数字的代码结构，其中前3位为主码，后4位为子码。主码的前两位数字用以惟一地定义要素的类别（如02表示地形要素；05表示水文要素；09表示边界要素等）；主码的第三位是子码的解释位：若为零，表示子码是要素的分类码；若非零，表示子码是要素的参数值或称为参数属性代码，如地形高程，路段长度等。子码的第一位数字通常为零，其余三位数字用以表示要素的图形类型（点状、线状或面状）、分级分类（计曲线、间曲线或助曲线等）和其他特征（洼地、冰川、水源或河流的左右岸等）。该编码标准逻辑性强、信息量丰富，便于进行要素之间关系的推理判别。

我国颁布的GB-2260-80编码标准对国内行政区的编码作如下规定：①码长最多10位；②省（市、自治区）占前3位，其余由用户自行定义。除此之外，测绘、国土、地矿等行业均根据国家要求和行业特点，制定了相应的分类编码标准。一般规定整形数4～8位，由高位到低位分别表示实体的类别、等级或某一特征。

此外，某些行业随着信息化过程的推进，已经将GIS作为其重要的技术手段和信息平台。人们在进行不同行业GIS开发和建设的同时，也加强了信息分类与编码技术的研究，有效地促进了行业GIS的发展。代表性的工作有中国煤矿信息分类与编码（吴立新等，1999；Wu L.X. et al.，1999a）、全球地学断面信息分类编码（许惠平等，2000）、内河电子江图信息分类编码（赵俊刚等，2001）和城市地下管网分类编码（王树东等，2001）。近年，国内有关行业针对自身行业特色和信息需求，在参考有关国际标准、国内规范的基础上，研究制定了自己的地学信息分类编码标准或体系，读者可以查看有关资料。

§7.2 分类编码模型

从认知地理现象、计算机存储地理数据、地理数据处理、空间查询分析，到可视化表达和最终地图制图，一般经过8个阶段，如图7-1所示（据Burrough P.A.，1987和焦健，1998改进）。其中过程（A）即对地理信息进行编码，是GIS设计中最重要的步骤，它代表现实世界与信息世界的界面（黄杏元，1989）。

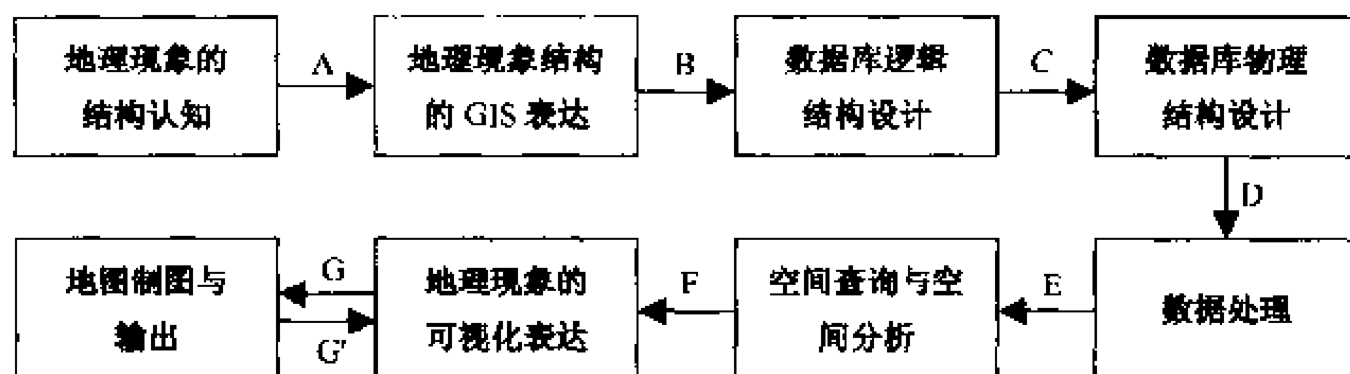


图7-1 从地理认知到地图制图的阶段性

将地理对象的定性和定量属性信息结合起来进行分类和分等定级，进而依类依级编码所构成的模型称为属性编码模型；将地理对象的空间位置和空间分布结合起来进行描述和定位，进而依类依位编码所构成的模型称为地址编码模型。

最典型的地理属性编码模型是树型结构模型和多维指标结构模型。如图7-2所示，树型结构模型是以属性分类等级序列为基础，每一层次的不同属性分别用一位或若干位代码（代码可以是数字型，也可以是字符型）进行表示，因而某一类别的地理对象的属性编码值是由其所有上级代码加本级代码顺序组成。对于同一专题内容，其代码的位数相对比较确定；对于不同的专题内容，其代码的位数可以相同也可以不同。而多维指标结构模型中的每一位代码分别表示一种属性。以土地利用为例，水稻田的树型结构编码可以是“123”，其中首位“1”为耕地代码；第二位“2”为水田代码；第三位“3”则为稻田代码。水稻田的多维指标结构编码的各位则可能是“水质类型”、“泥土类型”、

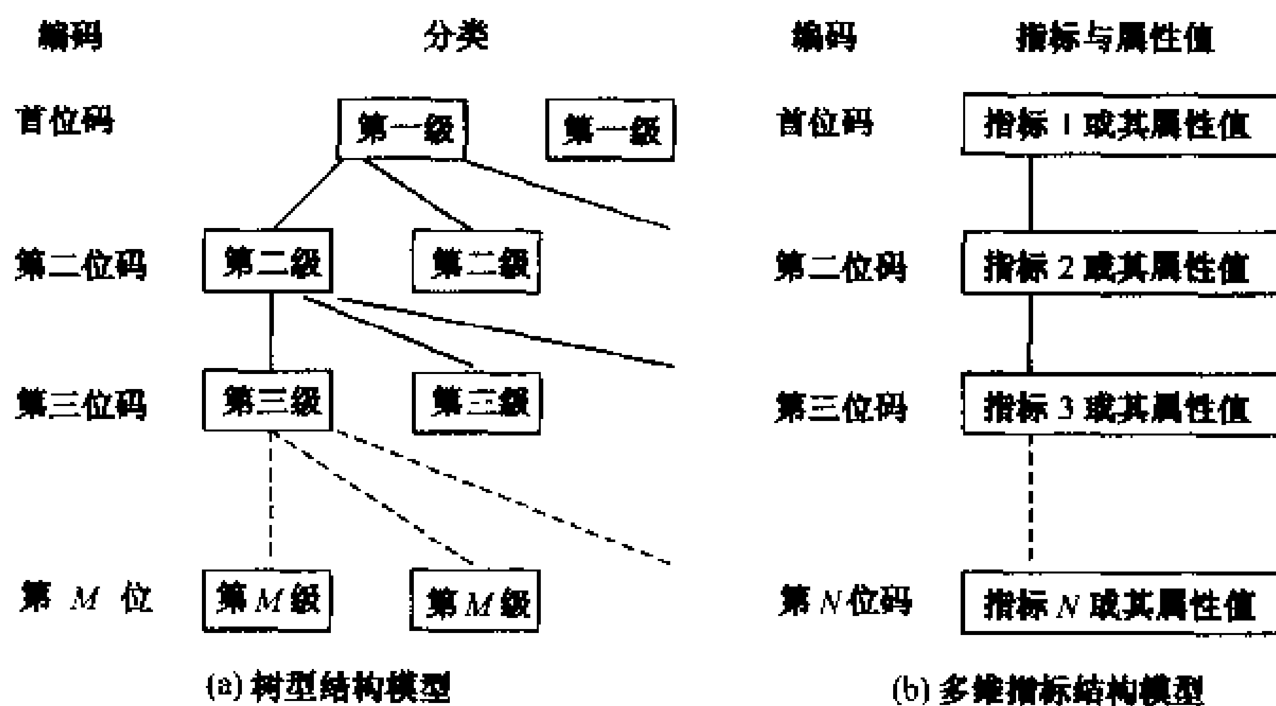


图7-2 地理属性编码模型

"pH" 等。根据专题研究对指标的需要，编码的位数可以扩展。

两种编码模型各有千秋，对 GIS 的数据采集与处理、空间分析与查询均有不同影响。其中，树型结构模型有利于地图图斑类别的确定；而多维指标结构模型则很容易派生出不同内容的专题图。

焦健等（1998）基于空间认知过程和地图模型的分析研究，提出地理编码可以区分为地理实体编码模型（Geographical Entity Encoding Model, GEEM）和图形实体编码模型（Cartographic Entity Encoding Model, CEEM）的观点。前者是用属性特征码、坐标及空间关系来描述和存储地理对象，包括树型结构模型和多维指标结构模型，是绝大多数 GIS 系统所采用的编码模型；后者是从地图的图形实体层进行编码，即用地图符号的图形属性码、坐标和符号之间的关系来描述地理对象，是地图生产者特有的模型，只有极少数制图软件（如美国的 AutoCAD 和中国的 MapCAD 等）采用。

如果从地理认知到地图制图的阶段性过程来看，GEEM 是发生在阶段（A）的正向行为过程；而 CEEM 则是发生在阶段(G')的逆向行为过程。两者各有利弊，前者易于派生新图和地图更新，便于地图形式的灵活和多样化表达，但编辑状态与制图状态之间的来回切换影响编辑查错和制图速度，降低了工作效率；后者则因编辑状态与制图状态统一（所见即所得），制图过程简单、直观、快速，但由于没有属性库或属性库很弱，难以进行属性查询和空间分析。

为了集中 GEEM 和 CEEM 在制图上的优势，焦健等（1998）提出基于两者混合的新的编码模型，即地理属性和图形属性耦合的信息编码模型（Coupled Geographical - Cartograpic Attribute Encoding Model, CGCAEM）。该模型同时用地理属性码和地图图形属性码来描述空间实体的属性信息，其逻辑构成如图 7-3 所示。其中地理属性信息编码仍然遵循地学分类系统拟定的编码原则，地图图形属性的编码位数基本上由地图符号的颜色、形状、线型、尺寸、角度、纹理等基本图形要素组成，编码位数固定不变。

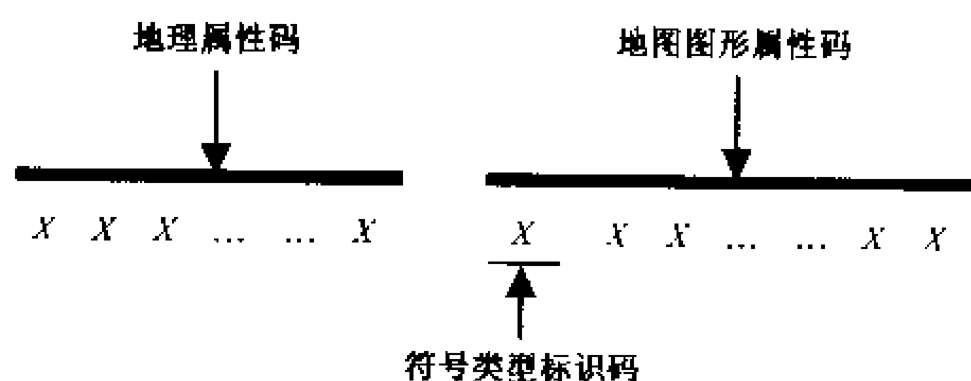


图 7-3 CGCAEM 码位的逻辑构成
（焦健等，1998）

§7.3 无边界 GIS 空间编码

7.3.1 传统的 ID 编码法分析

前已述及，通常非对象型 GIS 中的属性数据和图形数据是分开管理的，即属性数据用关系型数据库进行管理，而图形数据用空间数据库进行管理。GIS 采用 ID 索引作为桥梁，连接地理对象的属性和图形数据，如图 7-4 所示。因而，ID 是 GIS 系统中最

重要的和最频繁使用的基本要素。ID 编码的机制直接影响空间数据库乃至整个 GIS 系统的效率和功能。因此，ID 必须具有持久的惟一性。

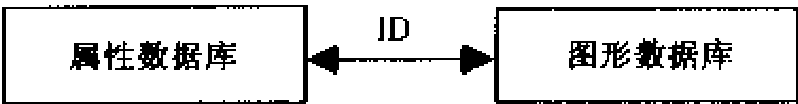


图 7-4 GIS 空间数据库组织连接

传统的 ID 编码技术包括命名法（denomination method）和分配法（collation method）。命名法虽然可以在静态和小区域条件下保证 ID 码的惟一性，但在大区域和动态条件下（如城市与矿区），由于空间实体的海量性和信息的动态性，ID 码的无错误命名及其惟一性均难以保障，即使通过努力完成了 ID 编码，码的长度将很长而不符合数据压缩的要求。分配法在单机环境和静态数据库条件下可能是有效的，但随着 Web-GIS 、分布式空间数据和多维动态 GIS 技术的发展，这一由机器随机分配 ID 码的方法已不再适用。

研究区域的不断扩大和持续发展，空间分辨率的不断提高，时空动态数据库建设的需要，均对传统的 ID 编码技术提出了严峻的挑战。由于命名法和分配法机制的限制，显然无法满足空间信息技术和多维动态 GIS 技术发展的需要，必须研究建立一种新的适合无边界 GIS 的 ID 编码新技术。

7.3.2 QuaPA 法编码原理

围绕矿山多维动态复杂 GIS 的理论与技术研究，作者（Wu L.X. et al., 1999b）曾研究提出一种适合矿区边界动态发展和环境动态变化的编码方法，即含象限前缀的主副码耦合编码方法，称为 QuaPA（Quadrant label prefixed for Principal code and Auxiliary code）法。QuaPA 法的实质是基于四叉树或八叉树的象限标识（反映空间实体所在宏观象限）与基于 Morton 编码的主码（反映空间实体具体位置）、基于传统编码的副码（反映局部区域内空间实体的惟一检索标识）的三位一体，分别服务于 2D GIS 索引和 3DGIS 索引。

1. QuaPA 编码规则

规则 1：空间实体的 ID 码由主码（principal ID, ID_1）和副码（auxiliary ID, ID_2）组成；

规则 2：主码代表空间实体的象限标识和轮廓位置，副码则用来在局域内惟一定义空间实体的名字和序列号；

规则 3：局域内实体之间的主码相同时，副码必须不同；不同局域的实体主码不同，其副码可以相同。

2. QuaPA 编码过程

第一步：首先按确定的单元尺寸将目标域（2D 或 3D）分割成 2D 矩形格网或 3D

空盒，然后由计算机自动计算确定每一格网或空盒的 Morton 标识码。该标识码即为格网内或空盒内所有空间实体的公共主码；

第二步：采用传统的 ID 编码技术，由用户对每一格网或空盒内所包含的空间实体进行逐一命名或分配序列号。所命名或序列号即为相应实体的独立副码。

其中副码的确定可以用前述的传统方法进行。如果是用命名法，则副码是明确的和清晰的，但编码效率较低，尤其是码的后期维护比较复杂；如果是用分配法，则编码过程可以由计算机来自动完成，但空间数据库的查询不太方便和直接。因此，两者各有优缺点，用户可以根据自己的习惯进行选择。

3．2D GIS 的主码自动编码

通常，地图的分幅号（map sheet numbering）可以直接作为 ID_1。但由于地图分幅将随地图比例尺而变化，因而可能导致以下结果：① 同一位置的块码在不同比例尺条件下可能不同；②同一实体的主码在不同比例尺条件下也可能不同；③不同比例尺条件下不同实体的主码可能一样。这些情况在空间数据库组织过程中是严格禁止的。因此，地图的分幅号不能直接用作 QuaPA 码的主码。

由于采矿活动是一个典型的时空动态过程，矿山 GIS（MGIS）所研究的对象和范围是随采矿进程而不断变化的。以某矿务局为例，该局由 5 年前的 5 个矿井发展到今年的 7 个矿井，3 号矿由 5 年前的一个开采水平（开采深度区间）、3 个采区（采矿活动平面区域）发展到今年的两个开采水平和 5 个采区；无论在平面区域还是深度范围上，该矿务局的 MGIS 的边界都发生了很大变化，而且还将随探明矿床范围和实际开采范围、开采深度的扩张而继续变化。因此，MGIS 是一类典型的无边界 GIS。矿山实体的编码是 MGIS 至关重要的内容，其生命周期横贯资源勘探、矿区开发、矿井生产和环境恢复全过程。

下面以矿山 GIS 为例，提出基于高斯坐标的 2D GIS 的 QuaPA 主码的自动编码步骤如下。

第一步：定义区域的平面参照基点 O（x_0，y_0）。参照基点应该是区域的稳定标志点。对于单一矿山而言，可以选择该矿主井的井口中心点；对于多个矿山或整个矿区，则可以考虑选择局办公楼的几何中心、旗杆点、某矿主井井口中心点或某一国家控制点。

第二步：定义沿 X、Y 轴的格网分割尺寸 D_x、D_y。结合四叉树原理，通常可以将 D_x、D_y 定义为 2^n（$n=1\sim10$）。其中 n 的选择是关键，依研究区域沿 X 、Y 轴的空间实体的数量、密度和复杂度而定。如果密度高且分布复杂，则 n 应取低值（$n=1\sim3$）；若密度较低且分布不太复杂，则 n 应取中值（$n=4\sim7$）；若密度很低且分布简单，则 n 可取高值（$n=8\sim10$）。

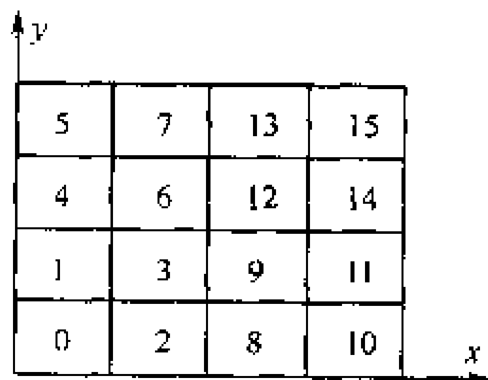


图 7-5 格网分割及格网编码

第三步：按格网分割尺寸 D_x、D_y 将目标区域进行格网分割。分割后的研究区内格网编码可以采用基于线性四叉树（LQT）的 Morton 编码方法，如图 7-5 所示，其计算方法见第四步。

第四步：确定任意空间实体 A 的主码。

（1）首先定义空间实体 A 的近基点：近基点是指实体 A 的包络矩形的最靠近基点 $O(x_0, y_0)$ 的角点，表述为 $A_0(x, y)$，具有 $\mathrm{Min}[(x-x_0)^2+(y-y_0)^2]$ 的特征；当存在两个近基点时，任选其中之一；

（2）然后计算 A_0 所在格网的十进制行号 I（沿 X 轴递增）和列号 J（沿 Y 轴递增）：

$$
\begin{aligned}
I &= \mathrm{int}\frac{|x-x_0|}{D_x} \\
J &= \mathrm{int}\frac{|y-y_0|}{D_y}
\end{aligned}
\tag{7-1}
$$

（3）将 A_0 所在格网的十进制行列号 (I, J) 转换为二进制行列号 (I', J')；

（4）获取二进制行列号 (I', J') 中每一位二进制数 P_{it}、P_{jt}：

$$
\begin{aligned}
P_{it} &= (I \& 2^{i-1}) \\
P_{jt} &= (I \& 2^{i-1})
\end{aligned}
\tag{7-2}
$$

式中：& 为“与”操作符；t 为位的序号，$t=1, 2, \cdots, n$；（若 P_{it}、P_{jt} 的位长不等时，应以较长者为参照对较短者补齐，即在较短者前面补零）。

（5）基于二进制行列号 (I', J') 计算实体 A 所在格网的二进制 Morton 码：

$$
I' = (i_n i_{n-1} \cdots i_1);
$$

$$
J' = (j_n j_{n-1} \cdots j_1)
$$

$$
\mathrm{MT} = i_n j_n i_{n-1} j_{n-1} \cdots i_1 j_1 \tag{7-3}
$$

（6）加入象限标识码，生成实体 A 的主码：定义分别以 $(00)_2$、$(01)_2$、$(10)_2$ 和 $(11)_2$ 作为第 0 象限 $((x-x_0)<0, (y-y_0)>0)$、第一象限 $((x-x_0)>0, (y-y_0)>0)$、第二象限 $((x-x_0)<0, (y-y_0)<0)$ 和第三象限 $((x-x_0)>0, (y-y_0)<0)$ 的象限标识。根据实体 A 的近基点 A_0 所在的象限位置确定其象限代码，并将该象限代码作为前缀加到实体 A 的 MT 码的首位，即得到实体 A 的主码 ID_1：

$$
\mathrm{ID}_1 = (\text{象限代码})\, i_n j_n i_{n-1} j_{n-1} \cdots i_1 j_1 \tag{7-4}
$$

实例：设已知目标区域的基点为 O(543721.73, 20732478.55)，实体 A 的近基点为 A_0(543440.45, 20733562.185)。定义：$D_x = D_y = 2^6 = 64\mathrm{m}$，那么：

$$
\begin{aligned}
I &= \mathrm{Int}(|543440.45 - 543721.73|/64) \\
&= \mathrm{Int}(|-281.28|/64) = \mathrm{Int}(4.395) = 4 = (00100)_2 \\
J &= \mathrm{Int}(|20733562.18 - 20732478.55|/64) \\
&= \mathrm{Int}(|1083.63|/64) = \mathrm{Int}(16.9317) = 16 = (10000)_2
\end{aligned}
$$

$$
\mathrm{MT} = (0100100000)_2 = (288)_{10}
$$

可知实体 A 属于第 0 象限，故实体的主码为：

$\mathrm{ID}_1 = (000100100000)_2$，

该主码占两个字节。

4. 3D GIS 的主码自动编码

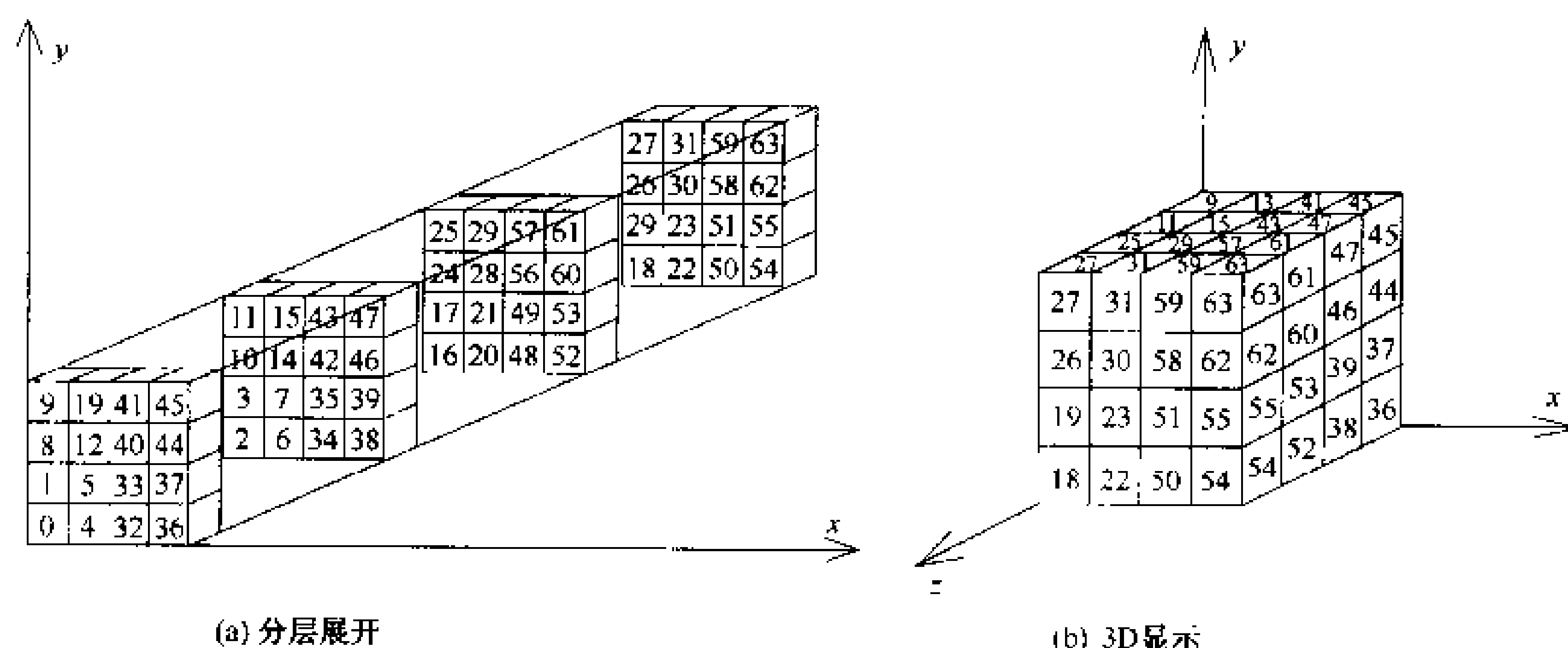
可以在 2D GIS 实体主码自动编码技术的基础上，将其推广到 3D GIS。3D GIS 的主

码自动编码步骤如下：

第一步：定义3D区域的基准点为 $O(x_0, y_0, z_0)$。对于单一矿山，可以定义为该矿主井井底车场的中心点；对于多矿或整个矿区，则可以考虑选择局办公楼的几何中心、旗杆点、某矿主井井口中心点或某一国家控制点。

第二步：定义沿 X、Y、Z 轴的格网分割尺寸 D_x、D_y、D_z。结合八叉树原理，通常 D_x、D_y 的定义同前，D_z 则定义 2^n（$n = 0 \sim 5$）。D_z 中 n 的选择根据沿垂直方向地层结构的复杂性和变异频度进行确定：当变异频度很高时，可选取低值（$n = 0 \sim 1$）；当变异频度不是很高时，可选取中值（$n = 2 \sim 3$）；当变异频度很低时，可选取高值（$n = 4 \sim 5$）。

第二步：将研究区域按尺寸 $(D_x \times D_y \times D_z)$ 分割为3D空盒集，如图7-6所示。此时，各空盒的编码规则与八叉树线形Morton编码原理一致，计算方法见第四步。



(a) 分层展开　　(b) 3D显示

图7-6　空盒分割及空盒编码示例

第四步：计算实体 A 的MT码：

(1) 首先定义实体 A 的近基点：实体 A 的近基点 $A_0(x, y, z)$ 为该实体的最小包络盒中离基准点最近的角点，特征为：$\mathrm{Min}[(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2]$。当存在两个近基点时，任选其中一个；

(2) 然后计算 A_0 所在空盒的十进制行号 I（沿 x 轴）、列号 J（沿 y 轴）和层号（沿 z 轴）：

$$
\begin{aligned}
I &= \mathrm{int}\frac{|x - x_0|}{D_x} \\
J &= \mathrm{int}\frac{|y - y_0|}{D_y} \\
K &= \mathrm{int}\frac{|z - z_0|}{D_z}
\end{aligned}
\tag{7-5}
$$

(3) 将 A_0 所在格网的十进制行列层号 (I, J, K) 转换为二进制行列层号 (I', J', K')；

(4) 获取二进制行列层号 (I', J', K') 中每一位二进制数 P_{it}、P_{jt}、P_{kt}：

$$
\begin{aligned}
P_{it} &= (I \& 2^{t-1}) \\
P_{jt} &= (J \& 2^{t-1}) \\
P_{kt} &= (K \& 2^{t-1})
\end{aligned}
\tag{7-6}
$$

式中：& 为"与"操作符；t 为位的序号，$t = 1, 2, \cdots, n$；(也需进行行列位长的补齐)

(5) 基于二进制行列号(I', J', K')计算实体 A 所在空盒的二进制 Morton 码：

$$
\begin{aligned}
I' &= (i_n \quad i_{n-1} \cdots i_1) \\
J' &= (j_n \quad j_{n-1} \cdots j_1) \\
K' &= (k_n \quad k_{n-1} \cdots k_1)
\end{aligned}
$$

$$
\mathrm{MT} = i_n j_n k_n \quad i_{n-1} j_{n-1} k_{n-1} \cdots i_1 j_1 k_1 \tag{7-7}
$$

(6) 加入象限标识，生成实体 A 的主码：按左手系定义分别以$(000)_2$、$(001)_2$、$(010)_2$、$(011)_2$、$(100)_2$、$(101)_2$、$(110)_2$ 和$(111)_2$ 分别作为第 0 象限[$(x - x_0) < 0$, $(y - y_0) < 0$, $(z - z_0) < 0$]，第 1 象限[$(x - x_0) < 0$, $(y - y_0) > 0$, $(z - z_0) < 0$]，第 2 象限[$(x - x_0) < 0$, $(y - y_0) < 0$, $(z - z_0) > 0$]，第 3 象限[$(x - x_0) < 0$, $(y - y_0) > 0$, $(z - z_0) > 0$]，第 4 象限[$(x - x_0) > 0$, $(y - y_0) < 0$, $(z - z_0) < 0$]，第 5 象限[$(x - x_0) > 0$, $(y - y_0) > 0$, $(z - z_0) < 0$]，第 6 象限[$(x - x_0) > 0$, $(y - y_0) < 0$, $(z - z_0) > 0$]和第 7 象限[$(x - x_0) > 0$, $(y - y_0) > 0$, $(z - z_0) > 0$]。

根据实体 A 的近基点 A_0 所在的象限位置确定其象限代码，并将该象限代码作为前缀加到实体 A 的 MT 码的首位，即得到实体 A 的主码 ID_1：

$$
\mathrm{ID}_1 = (\text{象限代码}) i_n j_n k_n \, i_{n-1} j_{n-1} k_{n-1} \cdots\cdots \, i_1 j_1 k_1 \tag{7-8}
$$

实例：设已知目标区域的基点为 $O(543721.73, 20732478.55, -200.0)$，实体 A 的近基点为 $A_0(543440.45, 20733562.185, -522.1)$。定义：$D_x = D_y = 2^6 = 64\mathrm{m}$, $D_z = 2^4 = 16\mathrm{m}$，那么：

$$
\begin{aligned}
I &= \mathrm{Int}(|543440.45 - 543721.73|/64) \\
&= \mathrm{Int}(|-281.28|/64) = \mathrm{Int}(4.395) = 4 = (00100)_2 \\
J &= \mathrm{Int}(|20733562.18 - 20732478.55|/64) \\
&= \mathrm{Int}(|1083.63|/64) = \mathrm{Int}(16.9317) = 16 = (10000)_2 \\
K &= \mathrm{Int}\,|-522.1 - (-200)|/16 \\
&= \mathrm{Int}\,|-322.1|/16 = \mathrm{Int}(20.13125) = 20 = (10100)_2 \\
\mathrm{MT} &= (011000101000000)_2 = (25216)_{10}
\end{aligned}
$$

可知实体 A 属于第 1 象限，故实体的主码为：

$\mathrm{ID}_1 = (001011000101000000)_2$，

该主码占 3 个字节。

7.3.3 QuaPA 法的全球推广

上节讨论的 QuaPA 编码法是以高斯坐标为前提的，其适用条件是不考虑地表曲率变化的小区域或高斯分带内部。当研究区域扩大到地表曲率不可忽略，或者研究区横跨多个高斯带乃至全球范围时，则存在高斯坐标与地球经纬坐标之间的不匹配或出现裂缝。

此时，需对 QuaPA 编码法做新的定义。

研究认为，通过对 QuaPA 法作一些补充约定，即可以将 QuaPA 法应用于全球空间数据集成组织，以解决全球空间实体编码的一致性和非重复性问题。其原理如图 7-7 所示。

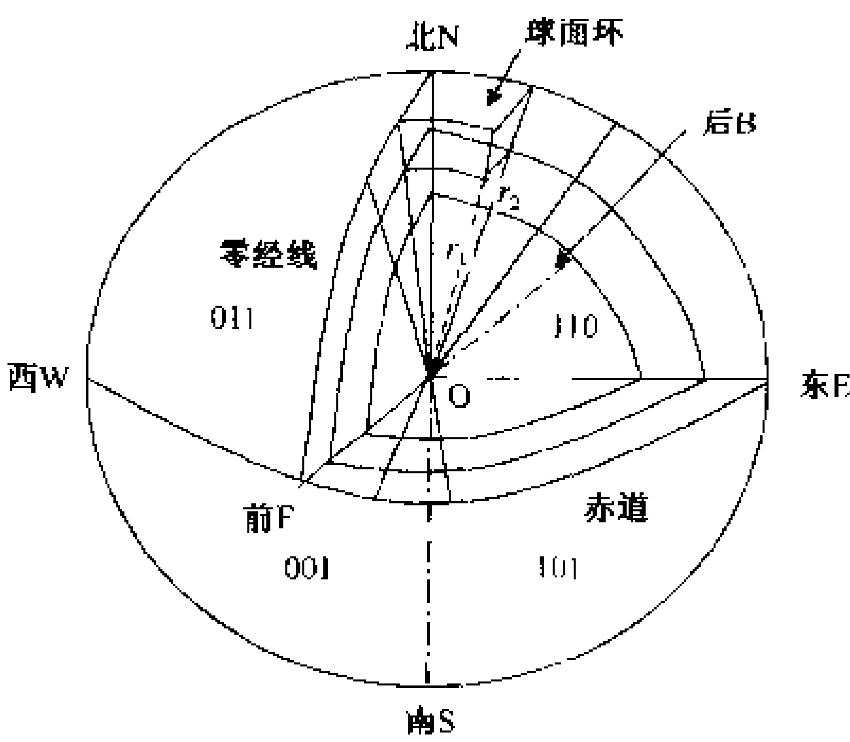


图 7-7　全球空间数据的 QuaPA 编码原理

定义 1：忽略地球椭球体长短轴的长度差异，将地球看成一个标准球体，按 WGS84 坐标系使用的椭球体长、短半径的均值，取其半径为：

$$R=\frac{a+b}{2}=\frac{6378137.0+6356752.3}{2}=6467444.6\ \text{m。}$$

定义 2：忽略地球重力异常，将大地水准面看成一个标准球面，其半径即为 R。

定义 3：以地心(如 WGS84 坐标系等地地球质心)作为全球 QuaPA 编码的基准点，其坐标为 O(0,0,0)；

定义 4：地球表面的 6 个特征点为东(E)、西(W)、南(S)、北(N)、前(F)和后(B)，其中南(S)、北(N)分别为南极点和北极点；东(E)、西(W)分别为西经 90°经线和东经 90°经线与赤道线(即零纬线)的交点；前(F)、后(B)分别为零经线(即过格林尼治天文台的经线)和 180°经线与赤道的交点。

定义 5：由分别经过南、北、前、后和东、西、前后的两个大圆将地球分割为 SBW、SFW、NBW、NFW、SBE、SFE、NBE 和 NFE 八瓣，相当于八个象限，分别以二进制(000)、(001)、(010)、(011)、(100)、(101)、(110)和(111)作为其标识码；

定义 6：以经圈、纬圈和半径为 $r(0\leqslant r\leqslant R)$ 的球面来分割地球，形成球面环层集，每个球面环层中的任一球面环即为 QuaPA 编码的一个子集，位于该子集中的空间实体具有相同的主码(包括象限标识码)；

定义 7：经圈、纬圈的分割尺寸 D_λ、D_φ 可以按 6°、3°或 1°等间距进行；球面分割尺寸 D_r 可以按等半径差进行，也可以不按等半径差进行(比如对于地球表层可以分割密一些，对于地球内部和地球外圈层可以分割的疏一些)。

定义 8：以各球面环中心的标称经纬度(λ',φ')和其上球面半径 r_i 作为参数，与基准点对应，计算球面环内空间实体的主码。标称经纬度(λ',φ')与实际经纬度(λ,φ)的对应

关系为：

$$\begin{cases} \varphi' = \varphi \\ \lambda' = \begin{cases} \lambda \\ \lambda - 90°(\text{当 } \lambda \geqslant 90°) \end{cases} \end{cases} \tag{7-9}$$

标称经度 λ' 在赤道平面的投影关系如图 7-8 所示。

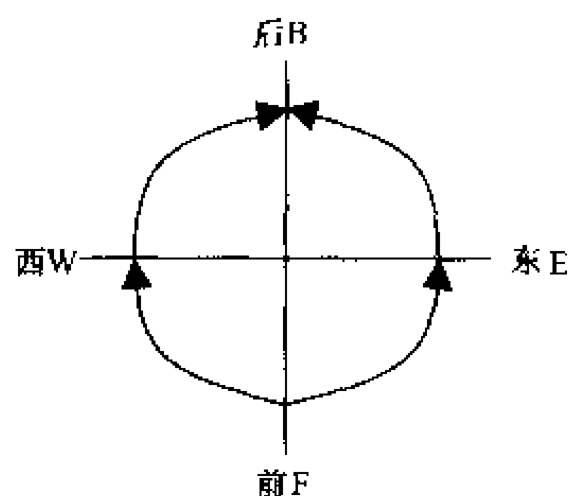


图 7-8 标称经度的度量

基于以上定义，面向全球空间数据组织的 QuaPA 码的主码的自动编码过程如下：

(1) 首先定义实体 A 的中心点：实体 A 的中心点 $A_0(\lambda, \varphi, r_1)$ 为该实体的最小包络环的几何中心点；

(2) 然后计算 A_0 所在球面环的十进制经圈号 I、纬圈号 J 和环层号 K（沿半径方向）：

$$I = \text{int}\frac{\lambda'}{D_\lambda}$$

$$J = \text{int}\frac{\varphi'}{D_\varphi} \tag{7-10}$$

$$K = \text{int}\frac{r'}{D_r}$$

(3) 将 A_0 所在格网的十进制经纬层号 (I, J, K) 转换为二进制经纬层号 (I', J', K')；

(4) 获取二进制经纬层号 (I', J', K') 中每一位二进制数 P_{it}、P_{jt}、P_{kt}：

$$P_{it} = (I \,\&\, 2^{t-1})$$

$$P_{jt} = (J \,\&\, 2^{t-1}) \tag{7-11}$$

$$P_{kt} = (K \,\&\, 2^{t-1})$$

式中，& 为"与"操作符；t 为位的序号，$t = 1, 2, \cdots, n$；（也有行列位长的补齐过程）

(5) 基于二进制纬经层号 (I', J', K') 计算实体 A 所在球面环的二进制 Morton 码：

$$I' = (i_n i_{n-1} \cdots i_1)$$

$$J' = (j_n j_{n-1} \cdots j_1)$$

$$K' = (k_n k_{n-1} \cdots k_1)$$

$$MT = i_n j_n k_n \, i_{n-1} j_{n-1} k_{n-1} \cdots \, i_1 j_1 k_1 \tag{7-12}$$

(6) 加入象限标识码，生成实体 A 的主码：根据定义 5，判断实体 A 的中心点 A_0 所在象限位置并确定其象限代码，并将该象限代码作为前缀加到实体 A 的 MT 码的首位，即得到实体 A 的主码 ID_1：

$$ID_1 = (\text{象限代码}) i_n j_n k_n \, i_{n-1} j_{n-1} k_{n-1} \cdots \, i_1 j_1 k_1 \tag{7-13}$$

实例：设实体 A 的中心点为 A_0(N54°34′40″, E20°33′56″, 6367522.1)。定义 $D_\varphi = D_\lambda = 3°$，$D_r = 2^8 = 256\text{m}$ 那么：

$$I = \text{Int}(54°34'40''/3)$$

$$= \text{Int}(54.51/3°) = \text{Int}\ (18.1) = 18 = (000000000010010)_2$$

$$J = \text{Int}(20°33'56''/3°)$$

$= \text{Int}(20.51/3) = \text{Int}(6.8) - 6 = (000000000000110)_2$

$K = \text{Int}(6367522.1 - 0)/256$

$= \text{Int}(6367522.1/256) = \text{Int}(24873.13) = 24873 = (110000100101001)_2$

$MT = (0010010000000000000001000000001100001010110001)_2$

可知实体 A 属于第 7 象限，故实体的主码为：

$ID_1 = (1110010010000000000000010000000011000010101110001)_2$

该主码占 6 个字节。

如果仅须对地球某一深度之内（如半径 6350000m 以上，约深 17500m 以内）的地物进行管理和数据组织，则可以该深度所对应的球面为基准面进行编码，并对 D_r 进行更细尺度的划分。若 D_φ、D_λ 尺度不变，$D_r = 2^4 = 16$m，此时实体 A 的中心点为 A_0 的经 纬 层号为：

$I = \text{Int}(54°34'40''/3) = \text{Int}(54.51/3°) = \text{Int}(18.1) = 18$

$= (00000010010)_2$

$J = \text{Int}(20°33'56''/3°) = \text{Int}(20.51/3) = \text{Int}(6.8) = 6$

$= (00000000110)_2$

$K = \text{Int}(6367522.1 - 6350000)/16 = \text{Int}(17522.1/16) = \text{Int}(1095.13) = 1095$

$= (10001011111)_2$

$MT = (0010000000000010001010010111111011)_2$

$ID_1 - (1110010000000000010001010010111111011)_2$

该主码占 5 个字节。

总之，空间数据编码，尤其是大型 GIS、全球型 GIS 或全球空间数据库编码，是一个极其重要而又复杂的课题。QuaPA 法为无边界 GIS 编码和全球空间数据编码提供了一种新的技术途径。QuaPA 编码用于无边界 GIS 和地球空间数据组织编码具有显著优势，包括：

1）与高斯坐标系和地理经纬系取得一致，容易与已有的地理、测绘成果结合；

2）在保证足够位置精度和较小分区（空盒或环层）的前提下，编码长度可以接受；

3）D_φ、D_λ、D_r 的选择有充分的自由性，不同尺度的选择为建立不同尺度（空间粒度）下的全球 QuaPA 码的多级体系结构提供了可能；

4）地球表面空间编码是地球整体编码的一部分，即为地球整体编码中的最外层编码，进而为地球表面空间信息组织提供了一致性框架；

5）为 GIS(Geographical IS)向 GeoIS(Geosciences IS)发展或 GIS 与 GeoIS 融合提供了统一空间编码参照。

此外，在空间索引方面，基于高斯坐标系的 QuaPA 编码法所定义的无边界 GIS 中的空间实体的近基点相当于窗坐标空间索引机制中的最小窗坐标，从而 QuaPA 编码法支持基于窗坐标的快速空间索引。由于 QuaPA 编码法基于对空间区域进行 2D 矩形、3D 六面体和球面环分割，QuaPA 编码法还支持格网索引机制。

参 考 文 献

国家技术监督局. 1986. GB7027-86 信息分类编码的基本原则和方法. 中国标准出版社，1～8

黄杏元. 1989. 地理信息系统概论. 高等教育出版社. 24－25
焦 健,曾琪明. 1998. 地理属性与图形属性耦合的信息编码模型. 遥感学报. 2(4):310～315
王树东,闫润海,程娟霞. 2001. 面向 GIS 的城市综合管网数据结构模式探讨. 测绘通报. 增:43－45
吴立新等. 1999. 矿业地理信息系统的信息分类编码研究. 见《1999'全国矿山测量学术会议论文集》:65－68
吴立新等. 2003 基于 QuaPA 原理的全球地学空间编码技术. 地理与地理信息科学. 2003, 19(5)(待发表)
许惠平等. 2000. 全球地学断面(GGT)GIS 的编码技术 长春科技大学学报 30(2):190－193
张振飞,马智民. 1999. 基于 GIS 的地学信息分类编码的一种思路. 地质科技情报. 18(3):95～98
赵俊刚,范世东,姚世玉南. 2001. 基于 GIS 的内河电子江图数据库信息分类编码研究及应用. 交通与计算机. 19(6):20－22
中华人民共和国民政部. 2001. GB2260—80:中华人民共和国行政区划代码. 见:《中华人民共和国行政区划简册》
Borrough P A, 1987. Principles of GIS for Land Resources Assessment
Wu L. X et al, 1999a. Information Classification & Management For MGIS And Digital Mine. In: Proc. of Int. Sym on DE Beijing, Science Pres. 999－1004
Wu L. X., Hu J. X., Xie C. L., 1999b. A ID-coding method containing quadrant label for non-boundary 3DGIS. In: Proc of Joint ISPRS Commission Workshop on Dynamic And Multi-Dimensional GIS (ISSN 0256-1840), Beijing, 105－108

第八章　地理元数据

随着计算机技术、网络技术及其他 GIS 相关技术的发展，使得空间数据共享共用及海量空间数据的集成与分布式管理成为可能与必然。然而，如何管理复杂的大型数据集，是数据生产者所面临的突出问题；同时，面对海量的地理空间数据，用户又往往无法知道哪些数据是他们真正需要的。

数据生产者需要有效的数据管理和维护办法，而用户则需要找到更快、更全面和更有效的方法，以便发现、访问、获取和使用现势性强、精度高、易管理和易访问的地理空间数据（蒋景瞳等，1999）。因此，关于空间数据的内容、质量、状态等的元数据信息显得更加重要，元数据已成为政府部门、非盈利组织、学术团体和企业共同关注的问题。不少国家和国际性组织已经发布和实施元数据内容标准，并开发了许多元数据操作工具。地理元数据标准和操作工具已成为国家空间数据基础设施（National Information Infrastructure, NII）的一个重要组成部分，是地理信息共享最为重要的标准之一，也是“数字地球”的重要支撑技术之一。

§8.1　信息共享与元数据

信息革命在全球范围的兴起，促使信息标准化工作日益升温。ISO 也意识到如果不依靠国际标准来引导信息化迅猛发展和信息量爆炸现象，必将出现一个信息混乱局面。鉴此，继采用 TCP/IP 标准解决 Internet 雏形——美国国防部“高级研究项目局”（Advanced Research Project Agency）的 ARPANET 网络的兼容问题之后，ISO 吸收 TCP/IP 的先进部分并完善其不足，提出了“开放系统互联参考模型（OSI/RM）”，以求得国际间信息通讯模式的统一。据统计，OSI 的正式标准目前已达 120 余项，国际上绝大多数厂商都向 OSI 体系过渡，各国政府（如美国、英国、欧洲等）也纷纷结合本国实际，以 OSI 为基本标准制定了相应的较精练的政府协议体系（Government OSI Profiles, GOSIP）。以上进展，为信息化社会的到来铺平了道路。

1. 信息共享研究的发展

信息化建设的中心问题是数据共享标准问题，其关键就是建立一套能够被大家普遍接受与采纳的标准，以确保信息的最大共享（承继成等，1998）。在计算机和空间技术冲击下的信息时代，地理信息被赋予了新的含义，其研究内涵已从传统的概念描述向机器语言表达和信息共享方面过渡，成为以数字地理信息为主的信息术语了。早在 ISO 做出反映之前，欧洲标准化组织（CEN）就已成立 TC287 技术委员会，致力于制定地理信息方面的数据标准，推出了 TC287 系列标准方案。鉴于 CEN 的先行工作，ISO 于 1994 年 3 月成立了“地理信息标准化技术委员会”，即 ISO/TC211，从事地理信息和地球信息业标准化方面的研究。该委员会与 CEN 达成了共同推进国际标准制定工作的协

议，使地理信息的标准化工作组正式纳入了国际标准化的轨道。我国于1994年加入该组织后称为P成员（积极成员），由国家技术监督局批准授权国家测绘局为ISO/TC211国内技术归口主管部门，目前已成立5个工作组开展包括元数据在内的地理信息共享标准研究。

2. 元数据的内涵

元数据的英文名称为Metadata，是关于数据的数据（data about data）。对于Metadata，汉语体系中有三种主要译法，即“元数据”（中国大陆）、“描述性数据”（中国大陆）和“诠释数据”（中国台北）。由于“描述”一词不能概括空间数据和关系数据的解释、说明、注释、控制等诸多功能，因而不尽贴切；“诠释”一词虽然有解释、说明和注释诸意，比“描述”的含义更广，但过于文学化，在科技、口语中不适用。“Meta”有词根的性质，词义包括中、亚、元、总、偏、超等，与IT相关的词汇中常译为“元”，如metacode（元代码）、metaprogram（元程序）、metacall（元调用）、metanotion（元概念）、metarule（元规则）、metasymbole（元符号）、metasystem（元系统）、metavariable（元变量）、metaword（元字符）等。因此，Metadata译为元数据比较贴切，也符合IT习惯。关于元数据及元数据字典问题，计算机领域的研究较多（秦晓等，1994）。

元数据到底为何物？实际上，图书馆的分类卡片、图书的版权说明页、CD唱片目录、磁盘标签、各类产品的用户手册和说明书等，就是一种元数据的具体表现。而在地学界，最典型的元数据便是各种纸质地图中的图例内容，包括图名、数据源、比例尺、精度、生产者、各类符号、生产日期或更新日期、注意事项以及可以在地图图廓上找到的标识信息均为元数据。通过元数据，人们拿到一张地图就便于理解和使用它。

当地图由纸质转变为GIS中的数字形式时，数据的管理和应用均产生一系列问题，比如：数据生产者需要在管理和维护海量空间数据的过程中不受工作人员变动的影响；用户如何以简洁、便捷的方式查找到自己所需的数据；以及得到数据后，如何正确无误地理解数据和转换数据。元数据可以帮助解决以上问题。目前的新型数字地理产品，如美国国防部基于DIGEST/VPF的Vmap™和DNC™等，均在数据集中载入了的元数据，从而保证信息的可持续管理与更新、便捷使用和最大共享。

3. 元数据的作用

元数据的主要作用可以归纳为以下几个方面（蒋景瞳等，1999）：

（1）帮助数据生产单位有效地管理和维护空间数据，建立数据文档，并保证即使主要工作人员退休或调离时，也不会失去对数据情况的了解；

（2）提供有关数据生产单位数据存储、数据分类、数据内容、数据质量、数据交换网络（Clearinghouse）及数据销售等方面的信息，便于用户查询检索地理空间数据；

（3）提供通过网络对数据进行查询检索的方法或途径，以及与数据交换和传输有关的辅助信息；

（4）帮助用户了解数据，以便就数据是否能满足其需求作出正确的判断；

（5）提供有关信息，以便用户处理和转换有用的数据。

4. 元数据的存在形式

元数据也是一种数据，在形式上与其他数据没有区别，可以以数据存储的任何一种方式存在，如卷宗式和数字式。

卷宗式是元数据的传统存在形式，可以以卡片、文件或手册的形式填写数据源和数据生产过程的主要信息，以便于阅读和数据查找，包括联机查询。

数字式元数据可以有多种方法建立、存储和使用。最基本的是文本文件，不受硬件和软件环境的限制，易于传输；另一种是用超文本建模语言（Hypertext Makeup Language, HTML）编写的超文本文件，用户可以利用 Netscape Navigator，Internet Explorer 或 Mosaic 进行浏览和查阅；第三种是用标准通用建模语言（Standard General Makeup Language, SGML）建立元数据及其间的连接，该法便于建立元数据索引和在空间数据交换网络上查询元数据，并且提供了一种在元数据用户之间交换元数据、元数据库和元数据工具的方法。

5. 元数据的可扩充性要求

元数据的可扩充性有两种含义，一是数据生产者对原有元数据的更新和扩充，此时只要在元数据表中添加相应元数据元组即可；二是由于元数据标准的变化和扩充，相应地要对已有元数据进行扩充，此时不仅会引入新的元数据内容，还可能引入新的结构。

当元数据存储采用文档方式存储时，由于元数据的存储不依赖于结构，故第一种扩充非常方便；第二种扩充若引入了新的结构，则与结构相关的部分可能受到影响，包括客户端界面和元数据文件的生成。

§8.2 地理元数据的构成与描述

8.2.1 地理元数据的构成

关于地理元数据，目前还没有统一的定义。欧共体的地理信息描述草案（CEN/TC 287）中，元数据被简单地定义为“关于数据的数据”；而在 ISO/TC211 中则被描述为“关于数据集的数据”。实际上，地理元数据是用于描述地理数据内容、定义、空间参照、质量及地理数据集管理等方面的数据，是地理信息标准化的重要组成部分，也是地理信息数据描述的关键技术之一。

地理数据通常组织成数据集，一个数据集包括一个或多个特征类型事件及它们的相关属性，也可以由其他数据集联合组成。元数据就是用来描述该数据集的基本特征和总体内容的数据信息，由元数据元素、元数据实体、复合实体、元数据子集 4 个层次构成，如图 8-1 所示。其基本单元是元数据元素，一个元数据元素说明地理空间数据某一方面的特征，相当于数据库的“字段”。一个或若干个元数据元素组成元数据实体；复合实体则由元数据实体、元数据元素和（或）其他复合实体构成；元数据子集是由若干元素、实体或复合实体组成的集合。

在元数据文本文件中，子集、实体和元素的关系，可以用逐级缩进的办法表示（称

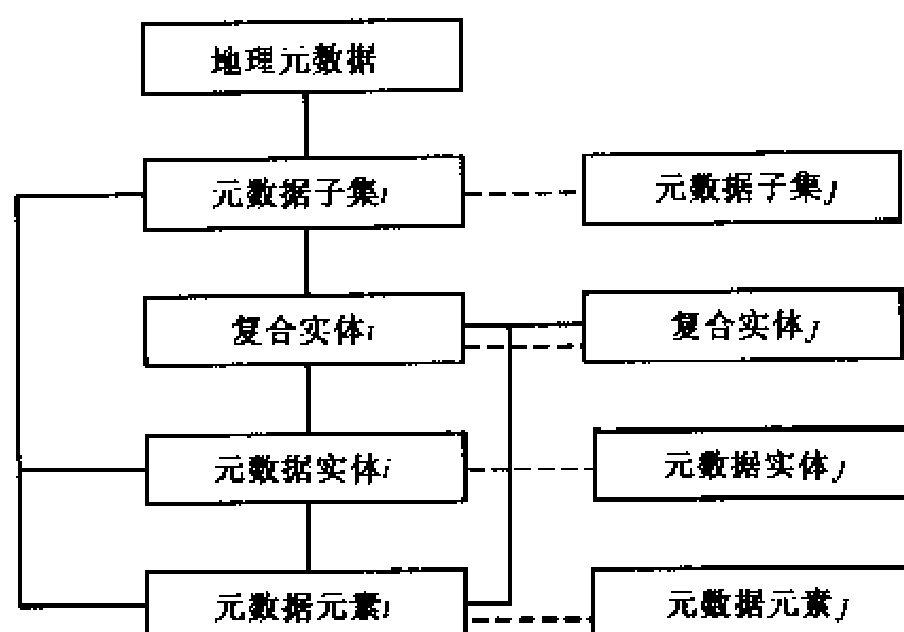


图 8-1 地理元数据的构成

为分层缩排)，或者用编号加小数点的方式表示（例如，实体 2.1 可以有元素 2.1.1 和 2.1.2 等)。美国地质调查局（USGS）发布的 30′×30′ DEM 元数据中的 Lineage 子集的以下部分内容，即体现了缩排结构（其中，黑体部分是元数据的数据值，其余是元数据的子集、实体和元素名称)：

Lineage:
  Source _ Information:
    Source _ Citiation:
      Citation _ Information:
        Originator: **U. S. Geological Survey**
        Publication _ Data:
        Title: **digital contour lines**
        Geospatial _ Data _ Presentation _ Form: **map**
        Publication _ Information:
          Publication _ Place: **Reston, VA**
          Publisher: **U. S. Geological Survey**
      Type _ of _ Source _ Media: **magnetic tape**
      Source _ Time _ Period _ of _ Content:
        Time _ Period _ Information:
          Range _ of _ Dates/Times:
            Beginning _ Date: **19880805**
            Ending-Data: **present**
        Source-Currentness _ Reference: **ground condition**

每个元数据元素、实体或复合实体均需要说明其名称、定义、类型、值域、简称等特征信息。子集、实体和元素可以具有 3 种性质，一种为“必需的”(Mandatory，简称 M)，即无条件必须提供的信息；第二种为“条件必需的”(Conditional，简称 C)，即一定条件下必需提供，如正在建立的元数据包含某子集、实体或元素，则必须提供其相应

的信息；第三种为“可选的”（Optional，简称O），即由用户决定是否将其包含在元数据文件中。

元数据一般分为两级。其中第一级包含最基本、最主要的实体和元素，即所谓的核心元素或编目信息，用于了解数据集的总体情况；第二级包含数据集的详细信息，往往可进一步分为若干子集，分别说明数据集某一方面的信息。

8.2.2 地理元数据的描述

通常，地理元数据通过以下9项内容进行描述（承继成等，1998），依序为：

（1）标识信息：简要描述数据集的信息，由数据集标题、可选标题（O）、缩写标题（O）等组成。

（2）数据集描述信息：全面描述数据的有关内容，包括摘要、目的、用途、几何子模式、空间参照系、参考文献（O）、示例（O）及相关数据集（O）等。

（3）数据质量信息：包括数据集的进程历史、专题精度、时态精度、完备性、逻辑一致性、点位精度等，是数据用户和数据生产者确定数据质量的必要信息。

（4）空间参照信息：包括直接空间坐标系和间接空间坐标系信息。其中直接空间参照系由水准、椭圆、地图投影及高程参照系等组成；间接空间参照系包括间接空间参照系类型和参照系日期（O）。

（5）时空信息：地理数据集在地理层和时态上不尽相同，应提供包括区域进程信息、数据集完备性、平面范围、垂直范围及时态范围等在内的较详细信息。

（6）空间数据表示信息：地理特征在不同的数据集中可以用不同的方法进行表示，因此对每一个数据集应分别提供特征类型描述、属性类型描述和相关类型描述（O）3类信息。其中特征类型描述由特征类型名称、特征类型定义、几何图元、结构图元、特征类型码（O）、事件（O）、特征专题精度（O）、点位精度（O）及完备性（O）组成；属性类型描述由属性类型名称、属性类型定义、属性类型码（O）、属性值域（O）、属性专题精度（O）及时态精度（O）组成；相关类型描述由相关类型名称、相关类型定义、起始特征类型、结束特征类型、基数、约束（O）、相关类型专题精度（O）及逻辑一致性（O）组成。

（7）分类信息：数据集的内容或专题范围可能通过其特征给出，也可能在分类系统中描述。如果在分类系统中描述，则特征会成为一级或多级结构的组成部分，并且当特征类型规定为一个等级分类时，应通过数据辞典的方式进行说明。分类信息由辞典（O）和辞典元素组成，其中辞典元素包括术语、定义、同义词、相关术语、广义术语、狭义术语，有时也通过图片（O）等加以说明；辞典则由辞典名称和辞典管理者组成。

（8）发行信息：为满足用户对所使用数据集的出处、使用权限及知情权，应提供生产或管理数据的组织信息、连接信息和数据发行信息。其中组织信息包括组织名称、组织地址、角色和组织功能（O）；连接信息包括连接点名称、地址及角色组成；分发信息包括约束、版权、价格政策、分配单位、媒介、格式、在线访问、订购及后援服务等。

（9）元数据参照系信息：包括元数据自身的日期信息、连接信息、标准信息、扩展

信息（O）、辅助信息（O）和约束信息等。

§8.3 地理元数据管理与操作

8.3.1 地理元数据管理

计算机中元数据的组织与编码是用软件实现的，其存储与管理可以是单一的数据文件（ASCII文件）或超文本文件等，也可以是关系型数据库。对于GIS和大型空间数据共享系统而言，其元数据管理的基本思路是：将地理元数据信息进行分类和规划，确定各元数据项的类型和长度，并建立相应的地理空间元数据库（Geospatial Meta-Database，GMDB），进而利于各种编辑工具开发地理元数据管理系统，实现对元数据信息的录入、浏览、查询、编辑、维护、检索、插入和删除等基本功能。元数据管理系统一般将元数据分为两类：文档元数据和数据字典。两者均属于元数据范畴，都是对空间数据集的描述，均包含实体和属性信息，这是它们相同的一方面；另一方面，两者在很多方面存在差异，如表8-1所示。

表8-1 文档元数据与数据字典的性质差异

（高登风等，1999）

方　面	文档元数据	数据字典
目的	作为文档提供给共享用户	提供定位数据集和图内查询所需信息
类型	文档型	控制型
内容	相对较多	相对较少
结构	可以没有	具有一定的结构
适用范围	使用数据集的任何用户	只在本系统内有意义

由于两类元数据的差别，系统对它们的管理方式截然不同。

1. 文档元数据的管理

通常GIS的属性数据是通过关系型数据库进行管理的，为了管理的统一化，可以将元数据也交于关系型数据库进行管理。但这样会带来一个问题，即元数据标准所规定的数据元素有着复杂的层次结构，而关系型数据库对层次结构的表达能力较弱。如此必然要采用很多的表来表达数据元素的层次结构，造成大量数据冗余和多表链接操作的麻烦。高登风等（1999）分析文档元数据的特点认为，由于文档元数据不参与查询操作，而只作为空间数据查询结果的文档之一提供给用户，因此文档元数据只有作为文档时才有意义。故可以在存储时打破元数据结构，在输出时再将各数据项进行重组，恢复其标准的结构，达到简化表结构、简化操作和降低数据冗余的目的。

2. 数据字典的管理

从计算机的角度，元数据字典（MetaData Dictionary，MDD）是关于数据字典

(Data Dictionary，DD)的DD，换言之，是关于DD的信息的集合。MDD中记录关于DD的以下内容：①DD的内外部结构（即内外部格式）及其转换关系；②DD数据的语义完整性信息；③其他信息，如关于DD的统计信息和状态信息等（秦晓，1994)。而从GIS的角度，数据字典的作用是描述空间数据在系统内的组织结构和空间数据集的部分特征，为查询过程中数据集的定位和图内实体-属性查询提供支持。在空间数据库系统的分析、设计、实现、调试、运行、维护和使用各阶段，全过程发挥作用（马智民等，2000)。因此，要求数据字典必须具有清晰的结构。

以规划建设中的国家地理空间数据交换中心（NGDC）的地理空间信息服务发布系统为例，其实质是一个将地理空间数据的生产者、管理者及其用户通过电子线路连接的分布式网络系统。该系统的元数据管理模型如图8-2所示。

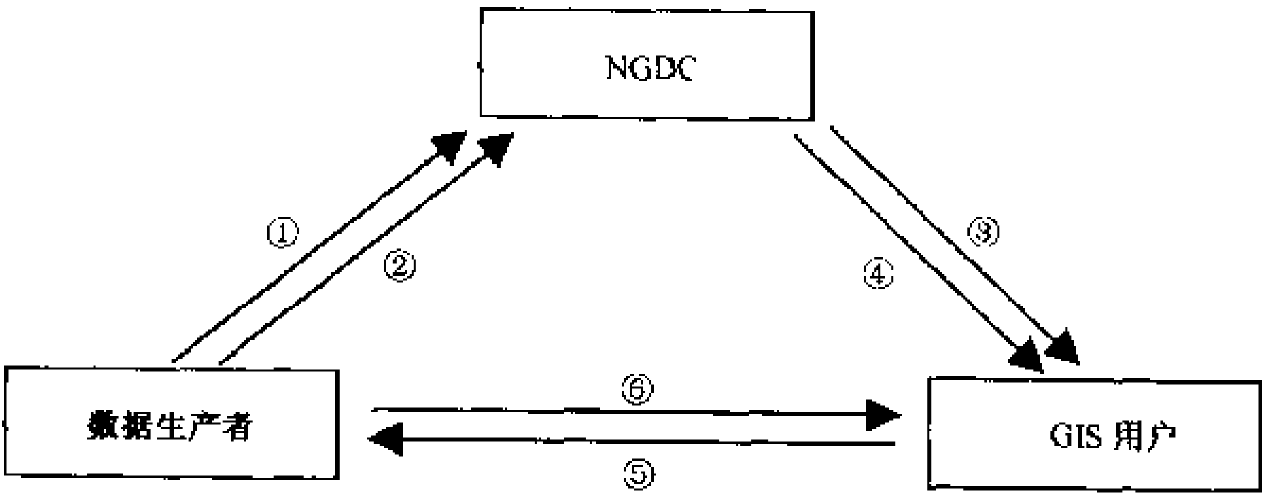


图8-2　NGDC中地理空间元数据管理模型

①提交地理空间数据及其元数据信息；②确认反馈；③发出元数据查询请求；④返回元数据查询结果；⑤发出地理空间数据产品需求请求；⑥发行地理空间数据产品（网络下载或邮寄）。

（据张立等，2000）

一般地，空间元数据管理系统可以采用客户/服务器（Client/Server，C/S）结构，并要求具有严格的与平台和操作系统的无关性。用户无论是PC机还是工作站，也无论是运行于Windows环境还是UNIX或Macintosh环境，都能使用该系统。基于分布式的空间数据库集成共享的空间数据交换中心的元数据服务器与分布式空间数据之间，可以采用星型结构实现网络连接，采用IE、Netscape或Communicator等通用浏览器来进行数据浏览。

8.3.2　地理元数据操作

文档元数据的操作包括创建、更新和输出，这些操作均由元数据管理子系统来实现。文档元数据的创建与更新可以由数据源客户端通过网络驱动，而输出则是由GIS用户端通过网络查询模块来驱动。元数据创建的同时要从元数据中抽取数据字典的内容，并将其添加到各数据字典表中，此后才可以完成空间数据的装入。元数据的更新是必须受权限控制的：一方面，对于中心数据库中的数据集，每个用户只能更新来自他自己的数据集的元数据；另一方面，元数据中的某些内容是不允许修改的。这种权限的检查以数据字典的内容为基础，只有权限被认可后才可以开始数据更新。查询模块在将查

询结果返回给客户端用户的同时，将内部格式的元数据按照元数据标准定义的结构重组，生成文件提供给用户浏览。

数据字典的操作主要是围绕数据抽取进行的。数据字典中的某些表并不是元数据的固有内容，只是系统对空间数据的某些特性进行抽取后的冗余存储，其目的是为了提高查询与图幅拼接效率。这些表包括图幅表、实体类型表、实体与图对照表（描述某个特定实体在哪些数据集的哪些图幅出现）和实体索引表（表达某个特定数据集中跨图幅的实体情况），其中前两个表由数据装入模块抽取，后两个表由元数据管理系统以图幅为单位进行抽取（高登凤等，1999）。当文档元数据发生更新后，可能会造成数据字典与空间数据集的不一致（跨图幅实体的增加或删除）。为了保持数据的一致性，必须在数据更新之后对两张表进行重新抽取，即重建实体与图对照表和实体索引表。

此外，由于元数据标准仅说明数据的内容而不是其格式，因而可能导致元数据即使符合标准，也可能因格式的差异而看起来不一样。这些格式差异包括：元数据的分层缩排、编号系统、元素名后加冒号、元素名称加下划线、单词分开或合并等。为了元数据管理方便和格式统一，可以用元数据操作工具来读入元数据文件，经过处理输出符合某一规范或习惯的元数据排列格式。

USGS 搜集整理了一套用于建立地理空间数据文档、符合 FGDC 元数据标准的元数据操作工具，如表 8-2 所示。这些工具可用于元数据的输入、编辑、预处理、后处理和确认工作。

表 8-2 元数据操作工具

（据蒋景曈，1999）

操作工具名称	功能	OS/GIS 平台
ASCII template document	帮助建立元数据条目	任何能处理文本的平台
BIC Metadata Form	在线元数据提交套件	客户端能处理表单的浏览器、服务器端能处理 CGI
Cns (Chew and Spit)	元数据编译预处理	UNIX, DOS, Win95 或 NT
Corpsmet 95	建立 CSDGM 元数据工具	Win95, NT
DATA DICTIONARY (AML)	Arc/Info 智能化元数据建立工具	UNIX ARC/Info
DataLogr 1.0	IDSN 元数据创建工具	MsWin 3.1、DOS
Dataset Catalog 4.0	元数据建立工具	有 Access 2.0 以上的 MsWin3.1, 3.11, Win95 或 NT
DOCUMENT (AML)	Arc/Info 智能化元数据条目工具	UNIX ARC/Info V7.0.3 或 7.0.4
FGDCNETA (AML) 1.1	Arc/Info 智能化元数据条目工具	UNIX 或 NT 上的 ARC/Info
The MDC (Metadata Collector)	元数据创建工具	MsWin, UNIX
Metadata Lite Entry Form	元数据提交套件	客户端用 Web 浏览器，服务器端用 Perl 和 Isite
Metadata Management System (LCRA)	CSDGM 元数据创建工具	MsWin 3.1, Win95, NT
Meta Data Manager professional 2.0	CSDGM、DIF、GILS 元数据工具	Win95、NT

续表

操作工具名称	功能	OS/GIS 平台
Metadata Validation Service	CSDGM 在线元数据编译器，用于实现元数据有效性检查	任何支持表单的浏览器
Metagen 32	CSDGM 元数据创建工具	Win95, NT
METALITE (AML) Beta 1.8	CSDGM 元数据创建工具	UNIX Arc/Info V7.0 以上
MetaMaker 2.10	CSDGM 元数据创建工具	MsWin
Mp	对 CSDGM 格式的元数据进行一致性检查，并输出为文本、HTML、SGML 或 DIF 格式	UNIX, DOS, Win95, NT
Xtme	CSDGM 元数据创建工具	UNIX with XR5

§8.4 地理元数据标准

标准化是人类生产与生活的重要组成部分，是人类文明的产物。其历史可以上溯至5000年前的埃及，那时该地区出现的可互换使用的弓与箭开始了古代标准化的萌芽。1900年前后，随着枪支的批量制造，从美国诞生了近代工业的标准化；1901年英国标准化学会的诞生，1906年国际电工委员会（IEC）、1926年国际标准化协会（ISA）和1947年国际标准化组织（ISO）的成立，为国际标准化的发展起了推动作用，使人类社会进入了一个崭新的时代，即现代化工业社会，也为工业社会向信息化社会的过渡奠定了基础。地理元数据标准是地理信息共享的最为重要的标准之一。该标准规定地理信息元数据的内容。包括数据的标识、内容、质量、状况及其他有关特征，可用于对各种地理信息的数据集的全面描述、数据集的编目及信息交换网络服务；实施对象可以是数据集、数据集系列、要素实体及属性。

8.4.1 国外与国际元数据标准

近年，许多国家、地区和相应的国际性、国家性专业组织在地理信息元数据标准制定方面做了大量工作，如美国联邦地理数据委员会（FGDC）、美国宇航局（NASA）和全球变化数据管理国际工作组、美国国际地球科学信息网络中心（CIESIN）、欧洲地图事务组织（MEGRIN）、加拿大通用标准委员会（CGSB）的地理信息专委会、中国国家基础地理信息中心（CNIC）等。以欧洲标准化组织（CEN/TC 287）、美国联邦地理数据委员会（Federal Geographical Data Committee, FGDC）和国际标准化组织（ISO/TC 211）为代表，所制定的地理元数据内容标准在国际上的影响最大，分别为 CEN/TC 287 地理信息元数据标准、地理空间数据元数据内容标准（CSDGM, 1995年开始执行第一版，1997年完成第二版）和地理信息元数据标准（ISO 15046-15，1994年启动研究，1998年提交各成员国审议）。

1. CEN/TC 287 元数据标准

早在 1992 年，CEN/TC 287 就开始了有关数字地理信息标准化方面的工作，并成立了 4 个工作组，从地理信息标准化框架（WG1）、地理信息模型和应用（WG2）、地理信息传输（WG3）、地理信息定位参考系统（WG4）等方面开始标准的制定工作，如表 8-3 所示。其目的是通过建立一系列结构化标准来建立一种用于定义、描述、传输和表现现实世界的方法，以促进地理数字信息的使用。其中，地理元数据标准的研究由 WG2 中的第 9 小组来执行。该小组所提交的地理信息元数据标准（CEN/TC 287）将元数据分为标识信息、数据集综述信息、数据集质量信息、空间参照信息、范围信息、数据定义、分类信息、管理信息、元数据参考及元数据语言等 10 部分，以此来描述数据集。

表 8-3　第二版 CSDGM 主要内容

工作组	标准
WG1：地理信息标准化框架	287001　参考模型 287002　综述 287003　定义 287005　数据描述技术 287013　查询与更新
WG2：地理信息模型和应用	287006　应用模式规则 287007　空间模式 287008　质量 287009　元数据
WG3：地理信息传输	287010　传输
WG4：地理信息定位参考系统	287011　位置 287014　间接定位系统

2. FGDC 元数据标准

FGDC 元数据标准 CSDGM 其实是一个参考文件，向用户说明数据获取、使用和评价过程中需要知道的事情，包括与元数据有关的术语和定义、说明哪些元数据是必需的、可选的或重复出现的、哪些数据是按 CSDGM 产生规则编码的，等等。其第二版 CSDGM（1997 年完成）包含 7 个主要子集和 3 个次要子集，如表 8-4 所示。

表 8-4　第二版 CSDGM 主要内容

主要子集	次要子集
标识信息	文献引用（引证）信息
数据质量信息	时间信息
空间数据组织信息	联系信息
空间参照系统信息	
实体和属性信息	
发行信息	
元数据参考信息	

FGDC的元数据标准自1995年开始在美国国内执行，加拿大、印度等国也已采用它作为自己的国家标准，ISO/TC211则基于该标准，研究制定相应的国际标准。

3．ISO/TC211元数据标准

ISO/TC211自1994年成立以来，就将"地理信息——元数据"列为其首批研制的20个国际标准之一，编号为ISO15046－15（ISO/TC211目前正在研制35个地理信息国际标准，其中元数据标准的编号为ISO19115）。经过几年努力，1998年5月完成的最后一版工作草案（WD V4.4）已提交ISO/TC211各成员团体征求意见。

该标准是以FGDC等现有标准为基础，按照国际标准化组织制定的标准规则要求制定的。该标准确定了两级元数据：①一级元数据：即编目信息，包含数据集编目所需的最少的元数据内容；②二级元数据：包含8个子集和3个可重复的实体。如表8-5所示。

表8-5　ISO/TC211二级元数据

8个子集	3个可重复实体
标识信息	文献引用信息实体
数据质量信息	负责单位信息实体
数据集继承信息	地址信息实体
空间数据表示信息	
空间参照系统信息	
应用要素分类信息	
发行信息	
元数据参考信息	

8个子集中，标识信息用来惟一标识一个数据集，说明其空间和时间范围、状况、法律限制和保密信息；质量信息描述数据质量的一般评价；数据志信息是有关数据集应用、数据源和数据集生产工艺的信息；空间数据表示信息是关于数据集中空间信息的表示方法；参照系统信息包括空间与时间参照；要素分类信息是关于数据集的要素类型、功能、属性和关系的定义与说明；发行信息是关于数据发行者及商业权限的信息；元数据参考信息是关于数据的现势性及数据负责单位。

3个可重复的实体中，文献引用信息实体是指提供引用文献的标准格式；负责单位信息实体是关于与数据集有关的单位和/或个人的标识；地址信息实体则提供与数据集有关的单位和/或个人的地址及其他通信办法。这3个可重复的实体不是独立的元数据子集，不能单独使用，而只是作为一个单元插在前述的某一子集中作为该子集的实体或元素。

该标准对每个元数据子集、实体及元素定义了8个特征，即名称、标识码、定义、性质、条件、最大出现次数、数据类型和值域。目前，该标准工作草案（WD V4.4）的一级元数据包含了55个元数据实体和元素、二级元数据共包含了452个实体和元素。

该标准是迄今最完整、也最为复杂的地理元数据标准方案。

ISO/TC211 的地理信息元数据的标准文本由标准化和信息化两部分组成，其中标准化部分是用户必须遵循的，而信息化部分则通过提供示例来帮助用户更好地理解标准。由于地理数据的多样性，难以保证某一元数据能满足所有不同用户的要求，故在标准化部分中还提供了一个可扩展方法，便于用户根据需求增加自己的元数据，并同时保证满足数据共享与互操作的需要。信息化部分则在分类系统中提供元数据及元数据在不同用户之间共享提供一个标准方法，同时也为元数据标准的使用提供应用指南和示例说明。

8.4.2 中国元数据标准

与国外相比，我国在元数据领域的研究、标准制定和实施方面均有较大差距（蒋景瞳，1999）。1997 年成立了全国地理信息标准化技术委员会（CSBTS/TC230），秘书处挂靠国家基础地理信息中心，负责我国地理信息国家标准的立项建议、组织协调、研究制定、审查上报等工作。近年，在国家有关部门的支持下，我国测绘遥感（蒋景瞳等，1999；刘若梅等，1999；张立等，2000）、地理地学（承继成等，1998；赵永平等，1998，1999；马智民等，2000；李军等，2001）、矿业、农业（范爱民等，2000；姚艳敏等，2001）、计算机（汪小林等，2001）等领域的有关单位和学者积极开展了地理信息元数据标准的研究，取得了一批有价值的研究成果，形成了一些标准雏形或提出了一些标准建议，并制定了一些与地理信息相关的国家标准（蒋景瞳，2002）。

1. 中国可持续发展信息共享元数据标准

在国家“九五”科技攻关项目的支持下，国家基础地理信息中心根据中国可持续发展信息共享示范系统总体设计的要求，考虑到参加共享的中央各部委的 60 个分布式异构数据库的内容和类型，参考 FGDC 的 CSDGM 和 ISO 的 ISO15046 - 15（V4.4）标准，研究制定了《中国可持续发展信息共享元数据标准》（方案 I）。

该标准方案由主体内容与适用范围、参考标准、术语、元数据分级、元数据内容、元数据扩展原则与方法共 6 部分内容（刘若梅等，1999）。该标准将元数据分为三层：即元数据子集、元数据实体和元数据元素；定义了元数据的 8 个特征：即名称、标识码、定义、性质、条件、最大出现次数、数据类型和值域；将元数据分为两级：一级元数据只包含编目信息，惟一标识一个数据集（包括数据集、数据集系列、要素和属性）所需要的最少的通用性元数据实体和元素（共 70 个），不再分子集；二级元数据为建立完整数据集（包括数据集、数据集系列、要素和属性）文档所需要的全部元数据实体和元素，与 ISO/TC211 标准一样，包括 8 个主要子集和 3 个可重复使用的子集（如表 8-4 所示），累计涉及 339 个实体和元素。

为便于管理、更新和查询，网络中心结点只建立一级元数据，而各部委分结点可同时建立两级元数据，或只建立二级元数据。用户可以从网络中心到分节点、由概括到详细、由简单到复杂循序渐进地了解数据集。按照标准规定的原则，用户可以扩展二级元数据，扩展类型包括：扩展元数据元素的值域、增加新的元数据元素、增加新的元数据实体类型、增加新的元数据子集、对已有的元数据元素增加更严格的限定和对已有的元

数据元素值域增加更多的限定。标准中定义了扩展原则和一般扩展方法（蒋景瞳，1999）。

2. 中国 NSII 元数据标准

通过对 CEN/TC 287、FGDC 和 ISO/TC211 元数据标准体系的分析，赵永平等（1998，1999）基于我国空间信息基础设施建设中元数据标准体系的急需，研究提出了一套中国 NSII 元数据标准体系的建议（方案Ⅱ）。该体系由 8 个标准化内容和 4 类在标准化信息中要被引用的内容，如表 8-6 所示。

表 8-6 NSII 元数据标准建议内容

（赵永平等，1998）

8 个标准化内容	4 个引用性内容
标识信息	引用信息
数据质量信息	时间范围信息
数据集继承信息	联系信息
空间数据表示信息	地址信息
空间参照系信息	
实体和属性信息	
发行信息	
元数据参考信息	

该标准将元数据分为三层：即部分、复合元素和数据元素；定义了元数据的 9 个属性：即编号、名称、英文名称、定义、约束条件、最大次数、数据类型、值域和标识码；将元数据分为两层：第一层为目录层，所提供的元数据复合元素和数据元素是从国家空间数据交换中心查询地理数据的目录信息，是第二层中一些必选项信息的概括，是相对宏观的信息；第二层是元数据标准的主体，包含如表 8-6 所示 8 个标准化内容和 4 个引用内容，即包括了全面描述地理空间信息的必选项、条件必选项和可选项。

3. 中国国土资源信息核心元数据

中国国土资源信息核心元数据的内容是在 ISO/TC 211 标准草案的基础上，结合中国国土资源空间与非空间信息的特点及描述要求而形成的。该核心元数据由 6 个标准化类和 2 个公共数据类构成，如表 8-7 所示，其数据字典如表 8-8 所示。

应该指出，随着分布式空间数据库技术的发展，越来越多的空间数据库开始与网络进行连接来提供数据共享。而已有的空间元数据标准的制定并没有很好地解决这一问题。汪小林等人（2001）的研究为此提供了一个有益的思路。他们在已有空间元数据标准的基础上，将空间元数据库分为全局级空间元数据库（Global Spatial Metadata Database）、区域级空间元数据库（District Spatial Metadata Database）和空间数据库级空间元数据库（General Spatial Metadata Database），并构造了一个基于网络的多层次空间元数据库框架，并以该框架为基础，提供了一种面向用户的简单数据请求描述方式，

进而给出了一个基于 Agent 的地理空间数据搜索获取策略。

表 8-7　中国国土资源核心元数据

（姚艳敏等，2001）

6 个标准类	2 个公共数据类
标识信息	覆盖范围信息
数据质量信息	负责单位联系信息
空间参照系统信息	
内容信息	
分发信息	
元数据参考信息	

表 8-8　中国国土资源元数据字典

（姚艳敏等，2001）

序号	名称	缩写名	定义	约束条件	最多出现次数	数据类型	域
1	MD_元数据	Metadata	关于元数据的信息	M	1	Class	行号 1～8
2	日期	MdTimeSt	元数据发布或最近更新日期	M	1	Class	CCYYMMDD (GB/T 7408－1994)
3	联系	MdContact	元数据负责单位的联系信息	M	1	Class	CI_负责单位《DataType》
4	标识信息	idInfo	数据集的基本信息	M	1	Association	MD_标识

参　考　文　献

承继成，赵永平．1998．地理信息及其元数据标准化．遥感学报．2（2）：149～154

范爱民．2000．矿区 GIS 建设中元数据问题研究．矿山测量．（1）：10～12

高登凤，杨冬青，唐世渭．1999．元数据管理与空间数据共享．见：蒋景瞳主编．中国地理信息元数据标准研究．科学出版社．29～35

蒋景瞳主编．1999．中国地理信息元数据标准研究．科学出版社．117～253

蒋景瞳，刘若梅，贾云鹏．1999．国际元数据标准的发展与研究现状．见：蒋景瞳主编．中国地理信息元数据标准研究．科学出版社．1～15

蒋景瞳．2002．地理信息标准化是实现信息共享的重要基础．地理信息世界．（2）：2～4

李　军．2001．地学数据集成的理论基础与集成体系．地理科学进展．20（2）：137～145

刘若梅，蒋景瞳，贾云鹏．1999．中国可持续发展信息共享元数据标准实施．见：蒋景瞳主编．中国地理信息元数据标准研究．科学出版社．36～46

马智民等．2000．多源地学信息系统数据字典设计与实现．西安工程学院学报．22（1）：13～17

秦　晓．1994．元数据字典及其实现．计算机学报．17（2）：81～87

汪小林等．2001．空间元数据研究及应用．计算机研究与发展．38（3）：321～327

姚艳敏，姜作勤，严泰来．2001．国土资源信息核心元数据的研究．测绘学报．30（4）：349～354

张 立，裴健雅．2000．地理空间元数据管理的研究与实现．武汉测绘科技大学学报．25（2）：127～131

赵永平，承继成．1998．地理空间元数据标准研究．信息技术标准化．（1）：8～11

赵永平等 1999．基于我国NSII关键技术研究的元数据标准内容体系．见：蒋景瞳主编．中国地理信息元数据标准研究．科学出版社．47～58

CEN/TC 287 WG2．1995．Geographical Information - Data Description - Metadata（Part I）

Digital geographical Information Working Group 1994．Digital Geographical Information Exchange Standards（Digest，V 1.2）

Federal Geographic Data Committee．1994．Content Standard for Geospatial Metadata．Washington DC

Metadata Working Group of FGDC．1998 Content Standard for Digital Geographical Metadata．http://www.fgdc.gov/

ISO/TC211 1996－12．Geographic Information / Geomatics 15046－15 Metadata（S）

ISO/TC211 WG3．1998．ISO Standards 15046－15．Geographic Information - Metadata，V4.4

ISO/TC211/WG3 CD15046－15．1998．Metadata

ISO/TC211．2000－06－01．Geographic Information Metadata（CD19115．3）．http：//www．statkart．no/

Tom H．1994 The Geographic Information System Standards Infrastructure．Standards View．（9）：133～142

第九章　地图投影与图形变换

GIS 表达和研究的主体是地球，而地球是一个不规则球体。如何将地球上各点投影在某一平面上进行连续的记录，是地图制图和 GIS 数据共享的基础。而在地图投影和 GIS 数据处理与可视化过程中，渗透着计算机图形学的基础理论与技术，其中图形变换已成为 GIS 基本原理的一个重要组成部分。

§9.1　地球与地理参数

9.1.1　地球的形状与大小

众所周知，地球是一个近似球体，其自然表面是一个极其复杂的不规则曲面，大陆的最高点（珠穆朗玛峰海拔 8848.13m）与海洋的最深点（马里亚纳海沟 - 11022 m）之间高差达 20 km。近年，国外学者根据人造卫星的轨道参数推算地球形状发现，地球是一个梨状体，北极位于梨柄处，南极位于梨底，若以赤道海平面到地心的半径作一圆，则北极海面高出此圆 418.9m，而南极海面则低于此圆 25.8m，如图 9-1 所示。

为了描述和表达地球表面，必须选择一个与地球形状、大小相接近的椭球体来近似代替它。大地测量中用水准测量方法得到的地面上各点的高程是依据一个理想的水准面来确定的，该水准面通常称为大地水准面。大地水准面是假定海水处于“完全”静止状态，将海水面延伸到大陆之下形成包围整个地球的连续表面；大地水准面所包围的球体称为大地球体。由于大地水准面上任何一点的铅垂线都与大地水准面成正交，而铅垂线的方向又受地球内部质量分布不均匀的影响而有微小变化，导致大地水准面产生微小的起伏。因此，大地球体仍然是一个表面起伏的不规则球体，还不能直接作为投影的依据。

为此，人们从数学角度出发，选择一个形状和大小与大地球体均极为接近的旋转椭球面来描述地球表面，称为地球椭球面，如图 9-2 所示。

地球椭球的两个主要参数为长轴半径 a、短轴半径 b，以及 3 个派生参数：

扁率：

$$\alpha = (a - b)/a \tag{9-1}$$

第一偏心率：

$$e = \sqrt{(a^2 - b^2)/a^2} \tag{9-2}$$

第二偏心率：

$$e' = \sqrt{(a^2 - b^2)/b^2} \tag{9-3}$$

一个多世纪以来，各国学者对地球椭球进行了众多研究，提出了多组地球椭球体参数，其中的常用参数数据如表 9-1 所示。

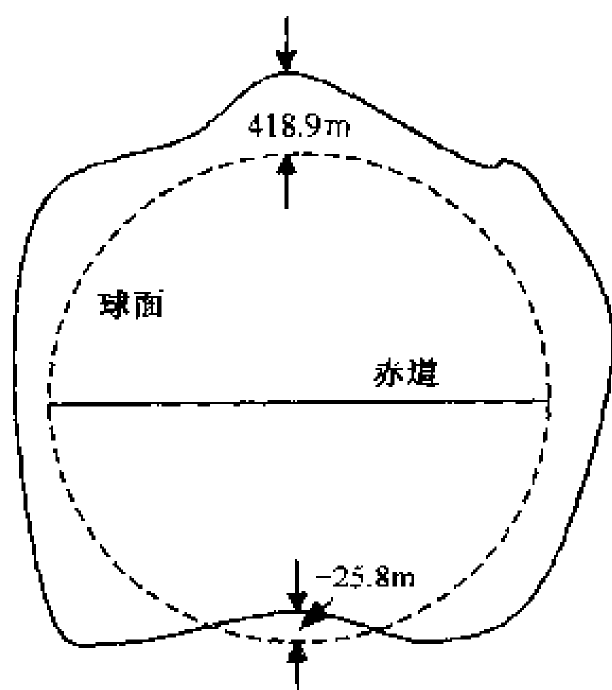


图 9-1　梨形地球

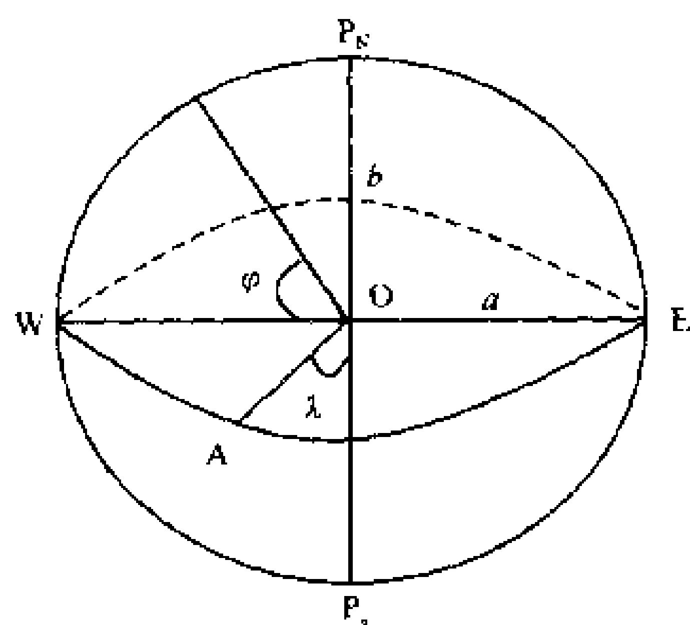


图 9-2　地球椭球体

表 9-1　常用的地球椭球体数据

椭球体名称	提出年代	长轴半径 a (m)	短轴半径 b (m)	扁率 α
Everest	1830	6 377 276	6 356 075	1:300.8
Bassel	1841	6 337 379	6 356 079	1:299.15
Clarke	1880	6 378 249	6 356 515	1:293.5
Clarke	1886	6 378 206	6 356 584	1:295.0
Hayford	1910	6 378 388	6 356 912	1:297
克拉索夫斯基	1940	6 378 245	6 356 863	1:298.3
IUGG	1976	6 378 160	6 356 775	1:298.25

我国在 1952 年以前采用 Hayford 椭球体；1953 年开始采用克拉索夫斯基椭球体；上世纪 70 年代末建立新的 80 坐标系时，采用 IUGG（国际大地测量与地球物理联合会）椭球体。而 1984 年定义的世界大地坐标系（WGS84）使用的椭球体长、短半径则分别为 6 378 137.0 m 和 6 356 752.3 m，扁率 $\alpha = 1:298.26$。

9.1.2　地理参数与坐标

地球是一个绕太阳公转的自转体，其自转轴线即图 9-2 所示的短轴，称为地轴。地轴与地面的交点 P_N、P_S 分别称为地球的北极和南极。

通过地心并垂直于地轴的平面称为赤道平面。赤道平面与地球表面的交线圆称为赤道，其半径即为地球椭球的长轴半径 a。平行于赤道的各个圆圈叫纬圈（纬线）、小圆或平行圈。

通过地轴并垂直于赤道面的平面叫做经面或子午面，它和地球表面相交的圈线叫经圈（经线）或子午圈。所有的经线长度均彼此相等，为一长轴半径等于 a 短轴半径等于 b 的椭圆。国际上统一以通过英国格林尼治天文台的经线作为起始经线，也称零经线。

由地球椭球体上任意一点M，可以引一垂线垂直于过该点的球面切平面（即地平面），该垂线也称为法线。法线与赤道面相交所构成的角度称为地理纬度，简称纬度，用 φ 表示。以赤道为起始，$\varphi=0$；往北称北纬，$\varphi=0\sim90°$（取正值）；往南称南纬，$\varphi=0\sim90°$（取负值）。

过地球椭球体上任意一点M的经线与零经线的夹角称为地理经度，简称经度，用 λ 表示。以零经线为起始，往东称东经，$\lambda=0\sim180°$（取正值）；往西称西经，$\lambda=0\sim180°$（取负值）。

经线和纬线是地球表面上两组正交的曲线，由其构成的坐标系称为地理坐标系。因此，地球表面任意一点M的地理位置，通常用其经纬度来表示，记为M（φ，λ）。地球表面两点经度或纬度之差，分别称为经差或纬差。如果两点位于同一经线上，则其经差为零；如果两点位于同一纬线上，则其纬差为零。

通过地心并过地面任意两点的平面与地球的交线，称为该两地面点的大圆。根据地球表面上两点的经、纬度，可以计算两点 M_1（φ_1，λ_1）、M_2（φ_2，λ_2）之间的大地线距离 D，该距离是两点大圆的弧长：

$$
\begin{aligned}
D &= R \cdot Z \\
Z &= \cos^{-1}[\sin\varphi_1\sin\varphi_2 + \cos\varphi_1\cos(\lambda_1 - \lambda_2)]
\end{aligned}
\tag{9-4}
$$

式中：R 为地球半径，一般取6371.11 km；Z 为两点之间大圆弧度值。

§9.2 地图投影的基本原理

数学中投影的含义是指建立两个点集之间一一对应的映射关系。在地图学中，地图投影是指建立地球表面上点（φ，λ）与投影平面上点（x，y）之间的一一对应关系。地图投影的目的是通过将不可展的球面投影到一个可展曲面上，然后将该曲面展开成为一个平面，来保证空间信息在地域上的连续性、完整性和可测度性。

9.2.1 地图投影的基本分类

据美国著名地图投影学家 J. P. Snyder 统计，全世界地图投影种类现有256种。依据不同目的和要求，可以采用不同的分类指标对如此繁多的地图投影进行分类。本节主要介绍四种主要的分类体系。

1. 基于投影面与球面相关位置的分类

通常，地图投影过程所采用的可展曲面有圆锥面、圆柱面和平面，分别对应圆锥投影、圆柱投影和方位投影。若以投影面与地球的相对位置进行区分，还可以进一步区分为正轴投影（投影面的中心轴与地轴重合）、斜轴投影（投影面的中心轴与地轴斜交）和横轴投影（投影面的中心轴与地轴垂直），如图9-3所示。由于各种投影均存在投影变形，距投影面越远，则变形越大。为控制投影的变形分布，可以调整投影面与地球椭球体的相交位置，根据这一位置的不同，可以进一步得到相应的切投影（投影面与椭球体相切）和割投影（投影面与椭球体相割），如图9-3中的1、5和9。

	正轴	斜轴	横轴
圆锥	① P P_1	② P P_1	③ P P_1
圆柱	④ P P_1	⑤ P P_1	⑥ P P_1
方位	⑦ P P_1	⑧ P P_1	⑨ P P_1

图 9-3　基于投影面与球面相关位置的基本分类

2. 基于投影方法的分类

根据投影获取的方法不同，可以将地图投影分为 3 类：

(1) 透视-几何投影：依据透视原理，根据视点（投影中心）、物点（地表的被投影点）与像点（投影平面上的点）之间的几何关系，来建立投影方程，如各种透视方位投影、空间投影等。

(2) 几何-解析投影：首先根据经纬线形状确定投影方程的基本形式，然后再依据某种约束条件解析地推求出特定投影的基本方程，如圆锥投影、圆柱投影和伪圆锥投影等。

(3) 解析投影：区别于几何-解析投影，并不事先设定经纬线的形状，而是完全依据人们给出的约束条件逐步推求经纬线的形状与投影方程，如高斯-克吕格投影、变比例尺投影等。

3. 基于投影方程的分类

这一分类完全依据地图投影方程中投影平面的直角坐标变量（X, Y）和投影平面的极坐标变量 ρ（纬线的投影半径）、δ（投影后经线与中央经线的夹角）与地球表面地理坐标（φ, λ）间的关系来划分。其分类结果如表 9-2 所示。

表 9-2　基于投影方程的分类

第一类	第二类	第三类
$\begin{cases}X=f_1(\varphi,\lambda)\\Y=f_2(\varphi,\lambda)\end{cases}$ 或	$\begin{cases}X=f_1(\varphi,\lambda)\\Y=f_2(\lambda)\end{cases}$ 或 $\begin{cases}\delta=f_1(\lambda)\\\rho=f_2(\varphi,\lambda)\end{cases}$	$\begin{cases}X=f_1(\varphi)\\Y=f_2(\lambda)\end{cases}$ 或
$\begin{cases}\delta=f_1(\varphi,\lambda)\\\rho=f_2(\varphi,\lambda)\end{cases}$	$\begin{cases}X=f_1(\varphi)\\Y=f_2(\varphi,\lambda)\end{cases}$ 或 $\begin{cases}\delta=f_1(\varphi,\lambda)\\\rho=f_2(\varphi)\end{cases}$	$\begin{cases}\delta=f_1(\lambda)\\\rho=f_2(\varphi)\end{cases}$

4. 基于投影变形的分类

根据地图投影可能引起的变形性质，可以将投影分为 3 种基本类型：

(1) 等角投影：投影后任意点处由任意两条微分线段构成的角度不产生变形。该种投影可以使得投影前后的形状保持不变，因而也称之为正形投影。

(2) 等面积投影：无论是微分面积还是区域面积在投影前后均保持不变。

(3) 任意投影：既不保持角度不变，也不保持面积不变，可能还存在长度变形。但当投影方向选择合适时，可以保证某一方向上的长度不变，此时称为等距离投影。

等角与等面积是相互抵消的，即等角是以牺牲等面积为代价的；而等面积则是以牺牲等角为代价的。任意投影虽然存在各种变形，但各种变形相对较均匀。

9.2.2　地图投影变形问题

由于地球表面是一个不可展为平面的曲面，在将这不可展的曲面转换为平面的过程中，必然引入投影变形。为了描述方便，引入长度、面积和角度 3 个量来衡量投影变形。

1. 长度变形与长度比

设地面上的一微分线段长度为 $\mathrm{d}l$，其投影后的长度为 $\mathrm{d}l'$，则投影后的长度变形绝对量 γ 和长度变形相对量即长度比 μ 分别为

$$\begin{aligned}\gamma &= \mathrm{d}l' - \mathrm{d}l\\ \mu &= \mathrm{d}l'/\mathrm{d}l\end{aligned} \tag{9-5}$$

当 $\gamma>0$ 或 $\mu>1$ 时，表示投影后长度增加；当 $\gamma<0$ 或 $\mu<1$ 时，表示投影后长度缩小；当 $\gamma=0$ 或 $\mu=1$ 时，表示投影后长度不变。

地球表面上任意一点沿经线或纬线方向的长度比分别称之为沿经线长度比 m 和沿纬线长度比 n，则任意点任意方向的投影后长度比 $\bar{\omega}$ 为：

$$\bar{\omega} = \sqrt{m^2\cos^2\alpha + n^2\sin^2\alpha + mn\cos\theta'\sin2\alpha} \tag{9-6}$$

式中：α 为任意方向的方位角，即过该点的大圆弧与该点所在经线的夹角；θ' 为投影后经纬线的夹角。

可见，投影后长度不仅是点位 (φ,λ) 的函数（影响 m，n 和 θ'），也是两点之间方位的函数。将一点上所有长度比中的最大值 (μ_1) 最小值 (μ_2) 称为极值长度比，

极值长度比所在的方向称为主方向。极值长度比可以按下式解算出来：

$$\begin{cases}\mu_1+\mu_2=\sqrt{m^2+2mn\sin\theta'+n^2}\\ \mu_1-\mu_2=\sqrt{m^2-2mn\sin\theta'+n^2}\end{cases} \tag{9-7}$$

2. 面积变形与面积比

设地面上的一微小面的面积为 da，其投影后的面积为 da'，则投影后的面积变形绝对量 V 和面积变形相对量即面积比 P 分别为

$$\begin{aligned}V&=\mathrm{d}a'-\mathrm{d}a\\ P&=\mathrm{d}a'/\mathrm{d}a=mn\sin\theta'=\mu_1\mu_2\end{aligned} \tag{9-8}$$

当 $V>0$ 或 $P>1$ 时，表示投影后面积增加；当 $V<0$ 或 $P<1$ 时，表示投影后面积缩小；当 $V=0$ 或 $P=1$ 时，表示投影后面积不变。

3. 角度变形

角度变形包括两个方面，即方位角变形和方向角变形。

椭球体表面某点相对于另一点的方位角定义为两点之间的大圆弧与过其中一点的经线之间的夹角。设投影前后椭球体表面某点相对于另一点的方位角分别为 α 和 α'，则 α' 可以用下式求得：

$$\alpha'=\mathrm{arccot}(m\cot\alpha+F) \tag{9-9}$$

方向角变形是指投影前后按主方向起算的角度差值，一般所指的角度变形是指此类。由于同一点处不同方向可能产生不同大小的角度变形，所以投影中常研究该点可能产生的最大角度变形 ω，其计算公式为：

$$\omega=2\arcsin\frac{\mu_1-\mu_2}{\mu_1+\mu_2} \tag{9-10}$$

§9.3 GIS 中地图投影选择

如图 9-4 所示，地图投影对 GIS 的影响渗透在 GIS 建设的各个阶段和各个方面。由

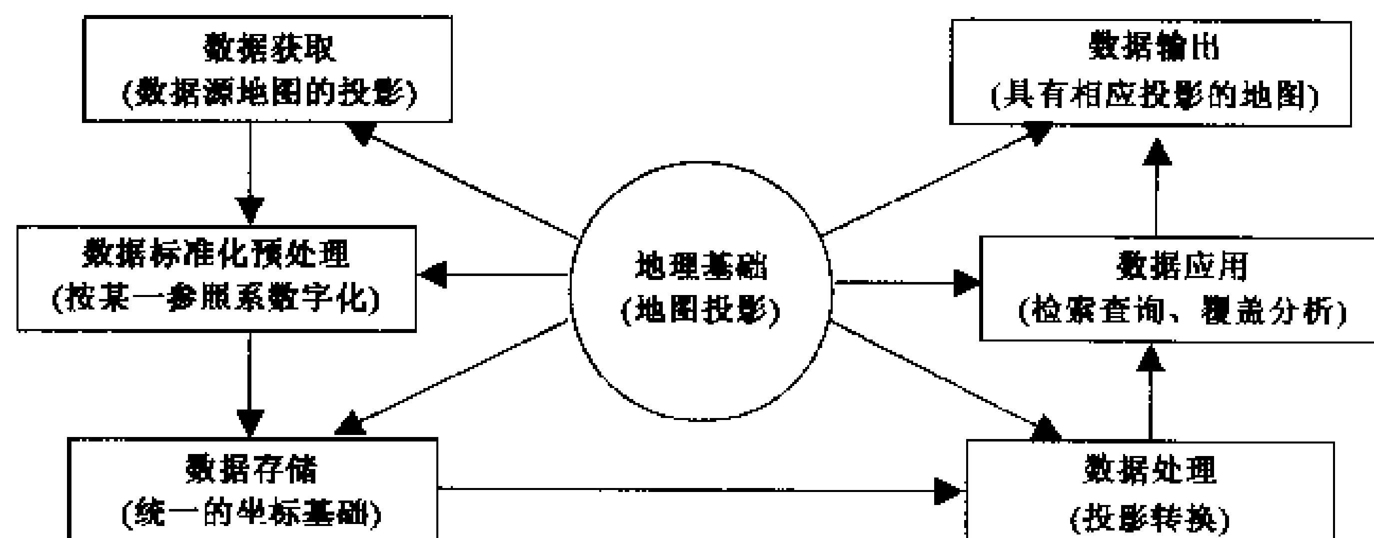


图 9-4 地图投影与 GIS 的关系
(据边馥苓，1996)

于不同的地图资料根据成图目的与当时需要而采用了不同的投影方式，而不同的投影方法具有不同性质和引入了不同的投影变形。当这些不同来源的地图资料进入 GIS 系统时，就存在一个统一的地理框架问题，即要用共同的地理坐标系统和网格坐标系统作为参照系统来记录存储各种信息，以此来保证同一系统中数据的交换、配准和共享。因此，在 GIS 的建立过程中，根据系统目的和地图比例尺的需要，选择合适的地图投影系统是必须首先考虑的问题。

9.3.1 地图投影选择的一般原则

统一的地理坐标系统是建立 GIS 的基础。国内外已经建立的各种 GIS 在选择地图投影系统时，一般遵循以下原则：

(1) GIS 所采用的投影系统应与本国的基本地图系列所采用的投影系统一致；

(2) 各比例尺 GIS 中的投影系统应与相应比例尺主要信息源地图的投影一致；

(3) 各地区的 GIS 投影系统应与该地区所使用的投影系统一致；

(4) 一般选择 1～3 种投影系统以保证地理定位框架的统一。

表 9-3 所示为加拿大、美国、日本 GIS 建设中所采用的地图投影方式与适用条件。

表 9-3 加拿大、美国、日本 GIS 投影选择

国家	比例尺	投影方式	备注
加拿大	1:2.5 万	UTM	地形图
	1:5 万	UTM	地形图、地质图、土地利用图、土地能力图
	1:12.5 万	UTM	地形图、土地利用图、植被图、
	1:25 万	UTM	地形图、地质图、土地利用图、植被图、土地能力图、自然资源图
	1:50 万	UTM	地形图、地质图、土地利用图、植被图
	1:100 万	Lambert	地质图、土地能力图
	1:110 万	Lambert	地形图
美国	州 GIS	州平面坐标系	基于 UTM，以 TM 和 Lambert 投影为主，局部地区采用了 HOM 投影
	≥1:50 万	UTM	国家基本地图系列
	< 1:50 万	地理经纬度	
日本		UTM	地形图、土地利用图、航空像片
		HOM	卫星像片

注：UTM 为通用横轴墨卡托投影；Lambert 为正轴等角割圆锥投影；HOM 为斜轴墨卡托投影；TM 为椭球体时的横轴墨卡托投影，即高斯投影。

9.3.2 中国GIS的地图投影选择

依照国际惯例并结合我国的具体实际，目前我国大多数已建、在建和待建的各种GIS均采用了与我国基本地图系列一致的地图投影系统，即大比例尺时的高斯－克吕格投影（横轴等角切（椭）圆柱投影）和中小比例尺时的Lambert投影（正轴等角割圆锥投影）。利用情况如下：

(1) 我国基本比例尺地图（1:100万、1:50万、1:25万、1:10万、1:5万、1:2.5万、1:1万、1:5000）除1:100万外均采用高斯-克吕格投影；

(2) 我国1:100万地形图采用Lambert投影，其分幅原则与国家地理学会规定的全球统一使用的国际百万分之一的地图投影保持一致；

(3) 我国大部分省区图以及大多数这一比例尺的地图也多采用Lambert投影或属于同一投影系统的Albers投影（正轴等面积割圆锥投影）；

(4) Lambert投影中，地球表面上两点之间的最短距离（即大圆航线）表现为近于直线，这有利于GIS的空间分析与度量。

1. 高斯-克吕格投影

高斯-克吕格投影的实质是将椭圆柱横切于地球椭球体上，其切线为某一经线（称为中央经线），然后根据一定的约束条件（即投影条件）将中央经线两侧规定范围内的点投影到椭球面上从而得到各点的高斯投影。

高斯投影条件为：

(1) 中央经线和地球赤道线投影成直线且作为该投影带（有限范围）的对称轴；

(2) 等角投影；

(3) 中央经线上没有长度变形。

根据上述条件推导出的高斯-克吕格投影的计算公式为：

$$\begin{cases} X = S + \dfrac{\lambda^2 N}{2}\sin\varphi\cos\varphi + \dfrac{\lambda^4 N}{24}\sin\varphi\cos^3\varphi(5 - \tan^2\varphi + 9\eta^2 + 4\eta^4) + \cdots \\ Y = \lambda N\cos\varphi + \dfrac{\lambda^3 N}{6}\cos^3\varphi(1 - \tan^2\varphi + \eta^2) + \dfrac{\lambda^5 N}{120}\cos^5\varphi(5 - 18\tan^2\varphi + \tan^4\varphi) + \cdots \end{cases} \tag{9-11}$$

式中：X，Y为点的平面直角坐标系的纵、横坐标；φ、λ为点的地理坐标，λ从中央经线起算，单位为弧度；S为由赤道至纬度φ处的子午线弧长；N为纬度φ处的卯酉圈曲率半径；$\eta = e'^2\cos2\varphi$，其中$e' = (a^2 - b^2)/b^2$为地球第二偏心率。

高斯投影的特点为：

(1) 没有角度变形，沿任意方向的长度比相等；

(2) 中央经线上无长度变形；

(3) 同一纬线上，离中央经线越远，变形越大；

(4) 同一经线上，纬度越低，变形越大；

(5) 等变形线为平行于中央经线的直线；

(6) 最大变形位于各投影带的赤道边缘处。

为了控制带内变形，我国地形图采用的分带办法是：1:2.5 万～1:50 万地形图采用 6°分带方案，即从格林尼治零经线开始起算，每 6°为一个投影带，全球共分为 60 个投影带，我国领土位于 72°～136°之间，共覆盖 11 个投影带（13～23 带），如图 9-5 所示。1:1 万及更大比例尺地形图采用 3°分带方案，全球共分为 120 个投影带，我国领土共覆盖 23 个投影带（24～46 带）。

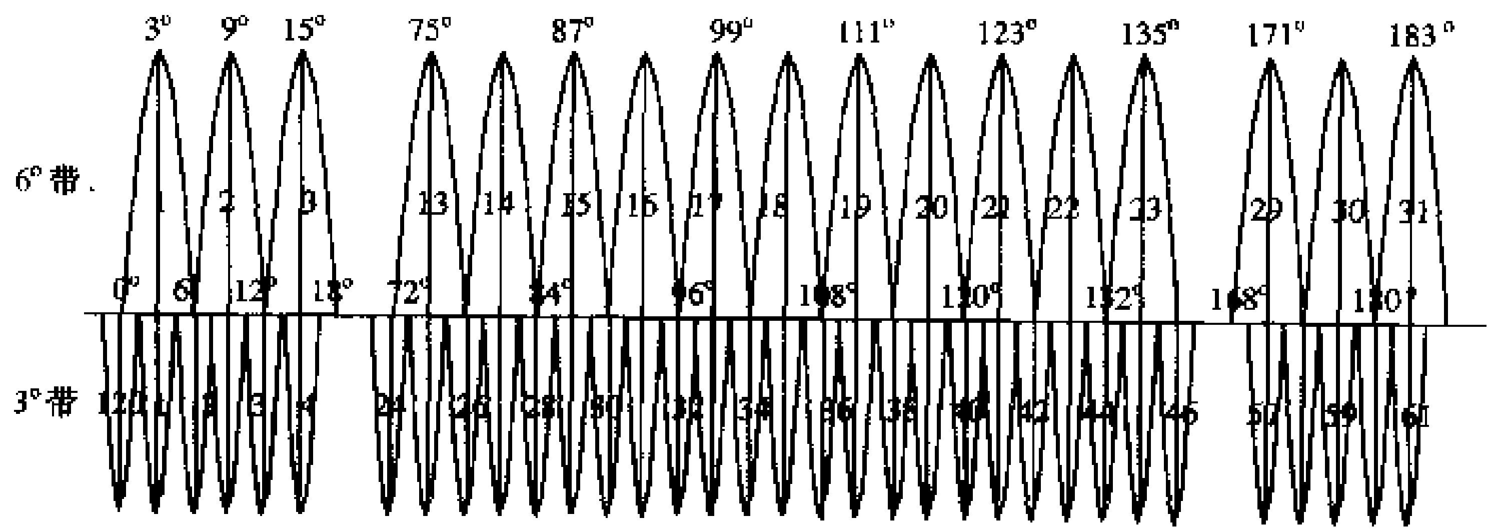


图 9-5　我国高斯投影分带情况

2．Lambert 投影

Lambert 投影是正轴圆锥投影的一个特例，其实质是用一标准圆锥面相割地套在地球（椭）球体上，按某种规则将地球表面上的要素投影到圆锥面上；然后将圆锥面沿某一母线（通常选择某一经线）展开，来得到地球表面的平面投影结果，如图 9-6 所示。

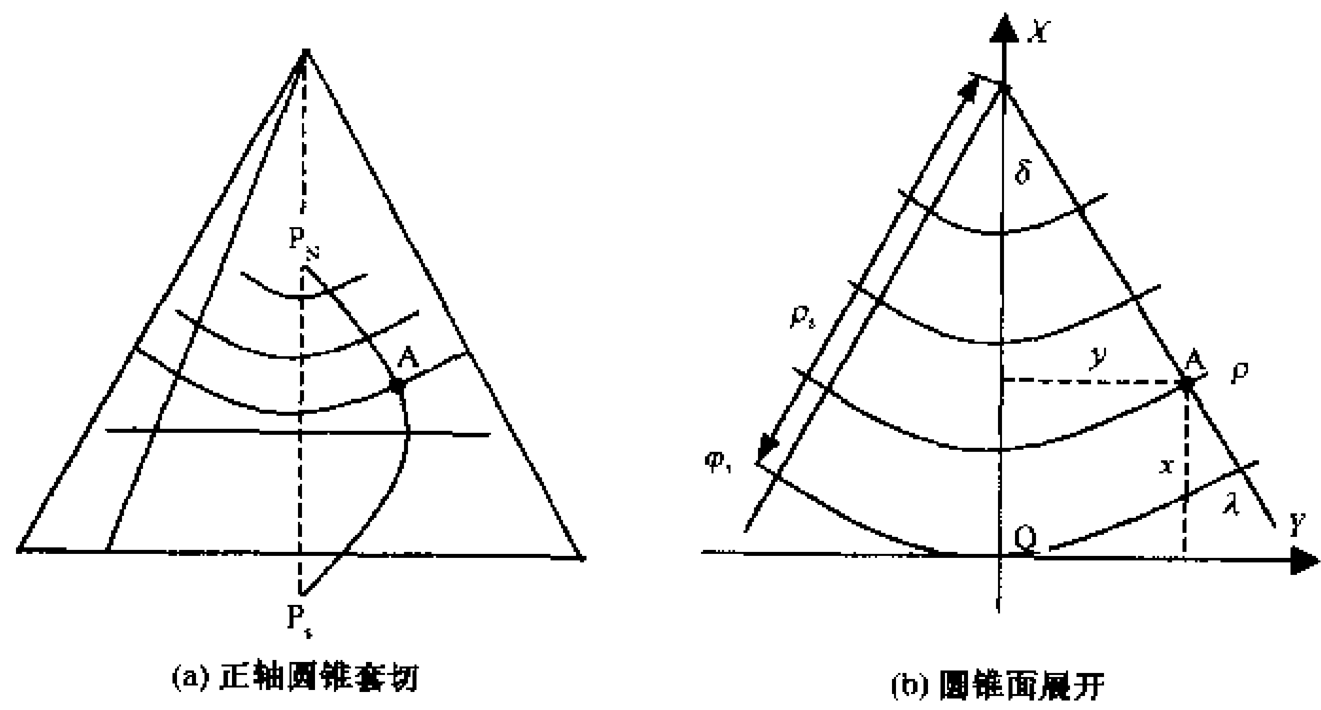


图 9-6　正轴圆锥投影原理

正轴圆锥投影的特点是：

(1) 投影后经线交于圆锥顶点；

(2) 纬线表现为同心圆，圆心为圆锥顶点。

正轴圆锥投影计算的一般公式可以写为：

$$
\begin{cases} \delta = \alpha\lambda \\ \rho = f(\varphi) \end{cases} \tag{9-12}
$$

$$
\begin{cases} x = \rho_s - \rho\cos\delta \\ y = \rho\sin\delta \end{cases} \tag{9-13}
$$

式中：δ 为投影后经线之间的夹角；ρ 为纬圈的投影半径；λ 为经差；α 为投影常数；ρ_s 为制图区域最低纬度的投影半径。

根据对正轴圆锥投影变形特征及其分布控制的要求，有等角投影、等面积投影和等距离投影 3 种典型模式，其投影方程与投影常数如表 9-4 所示。

表 9-4　正轴圆锥投影方程与投影常数

投影模式	投影方程	投影常数	
		单标准纬线（切投影）	双标准纬线（割投影）
等角投影	$\rho = k/U^{\alpha}$ $U = \frac{\tan(45^\circ+\varphi/2)}{\tan^{e}(45^\circ+\varphi/2)}$ $\sin\varphi = e\sin\varphi$	$\alpha = \sin\varphi_0$ $K = \frac{r_0 U_0^{\alpha}}{\alpha}$	$\alpha = \frac{\lg r_2 - \lg r_1}{\lg U_1 - \lg U_2}$　（Lambert 投影） $K - r_1 U_1^{\alpha}/\alpha = r_2 U_2^{\alpha}/\alpha$
等面积投影	$\rho^2 = \frac{2}{\alpha}(C - S_{\varphi})$ $S_{\varphi} = \int_0^{\varphi} Mr\mathrm{d}\varphi$	$\alpha = \sin\varphi_0$ $C = \frac{\alpha\rho_0^2}{2\alpha} + S_{\varphi_0}$	$\alpha = \frac{r_1^2 - r_2^2}{2(S_2 - S_1)}$ $C = \frac{\alpha\rho_1^2}{2\alpha} + S_{\varphi_1} = \frac{\alpha\rho_2^2}{2\alpha} + S_{\varphi_2}$
等距离投影	$\rho = C - S_{\varphi}$ $S_{\varphi} = \int_0^{\varphi} M\mathrm{d}\varphi$	$\alpha - \sin\varphi_0$ $C = N_{\varphi_0}\cot\varphi_0 + S_{\varphi_0}$	$\alpha = \frac{r_{\varphi_1}}{C - S_{\varphi_1}} = \frac{r_{\varphi_2}}{C - S_{\varphi_2}}$ $C - \frac{S_{\varphi_2} r_{\varphi_2} - S_{\varphi_1} r_{\varphi_1}}{r_{\varphi_1} - r_{\varphi_2}}$
备注	φ_0 为切投影时所切纬线的纬度；φ_1、φ_2 为割投影时所割纬线的纬度		

我国 GIS 中配置高斯投影和 Lambert 投影并不意味 GIS 仅处理这两种投影所提供的数据，或仅以这两种投影作为输出模式。实际上，几乎所有的商业 GIS 软件系统均提供了不同投影模式之间的数据转换（详见第 17 章）模块，GIS 可以接受各类投影方式的地图数据，并且可以按任意的投影方式进行数据和地图输出。配置高斯投影和 Lambert 投影的目的，主要是统一存储参照，以及统一分析和运算的地理基础。

9.3.3　面向数字地球的投影问题

“受传统地图空间数学基础理论和方法的影响，目前几乎所有的 GIS 均沿用地图投影作为自己参考系的数学基础，这在特定用途、局部范围内是可行的。”（胡鹏等，2001）但全球空间信息化进程的快速推进和数字地球战略的实施，在给 GIS 带来巨大

应用与发展空间的同时，也提出了一些新的问题。比如，在地图投影方面，一方面为了全球空间数据共享和大型 GIS 建设，需要一套新的实用的“地图投影”模型，来满足全球多分辨率连续可视化和精确量算与地理分析的需要；另一方面，为了强调区域特征或灵活方便地表达城市地图，需要编制变比例尺的地图，来满足人们对重点区域的细节信息可视表达的需要。

尤其在全球空间数据共享和地球三维连续表达方面，人们最为关注。全球空间数据基础设施（GSDI）委员会及技术工作组主席 Peter H.（1999）在北京“数字地球”国际会议上的报告中曾指出，该技术组将采用地理坐标作为空间参考体的惟一标准，来作为国际标准和在过渡期内与各国标准相关联，以此来支持数字地球的需要。陈俊勇院士（2000）也提出，大地测量要由过去的二维平面系统向高精度、三维、动态的坐标框架转变。

近年来，人们对上述问题进行了多方面的研究，取得了一些初步成果，包括由传统的曲面到平面的投影变换扩展为曲面到曲面的投影变换（胡鹏等，2001），“放大镜”式方位地图投影（Snyder J. P.，1987）、组合方位投影（杨启和，1988）和组合投影（杨晓梅等，1998）等变比例尺地图投影模式。

§9.4 图形变换

计算机图形学是近 30 年来随着计算机及其外围设备的发展而迅速发展起来的一门新型学科，主要研究用计算机及其图形设备来输入、表示、变换、运算和输出图形的原理、算法及系统。目前，计算机图形学已经在用户接口、计算机辅助设计与制造（CAD/CAM）、交互绘图、勘探测量、过程控制与环境模拟、电子印刷与办公自动化、艺术模拟、工业模拟、科学计算可视化及 GIS 等领域得到充分的应用。在 GIS 方面，计算机图形学发展为 GIS 的图形变换提供了坚实的图形学基础；而 GIS 的可视化需求也反过来促进了计算机图形学中图形变换理论与算法的丰富和发展。

9.4.1 窗口区和视图区的坐标变换

通常，用户指定用户域（WC）中的某一区域为工作窗口（W），W 小于或等于 WC，任何小于 WC 的 W 均称为 WC 的一个子域。窗口通常是矩形的，可以用其左下角和右上角的坐标来表示，或用其边长来表示；窗口可以嵌套，即可以在第 i 层窗口中再定义第 $i+1$ 层窗口。屏幕域是设备输出图形的最大区域，是有限的整数域，如对于分辨率为 1024×1024 的显示器，其屏幕区域为 $DC \in [0:1023] \times [0:1023]$。任何小于或等于屏幕域的区域称为视图区，可以由用户在屏幕域中用设备坐标来定义；在一个屏幕上，可以定义多个视图区，分别作不同的应用，如主菜单区、图形显示区、提示信息区等。

在图形显示区，必须进行从窗口区到视图区的坐标变换，即将窗口区内景物的实际坐标变换为视图区的屏幕坐标，亦即像素行列位置。

1. 变换公式

如图 9-7 所示，用户坐标系下窗口区的左下、右上角的坐标分别为（W_{XL}，W_{YB}）、（W_{XR}，W_{YT}）；其相应的屏幕视图区的左下、右上角的坐标分别为（V_{XL}，V_{YB}）、（V_{XR}，V_{YT}）。则窗口中的任意点（X_W，Y_W）对于屏幕视图区中的点（X_V，Y_V），其变换公式为：

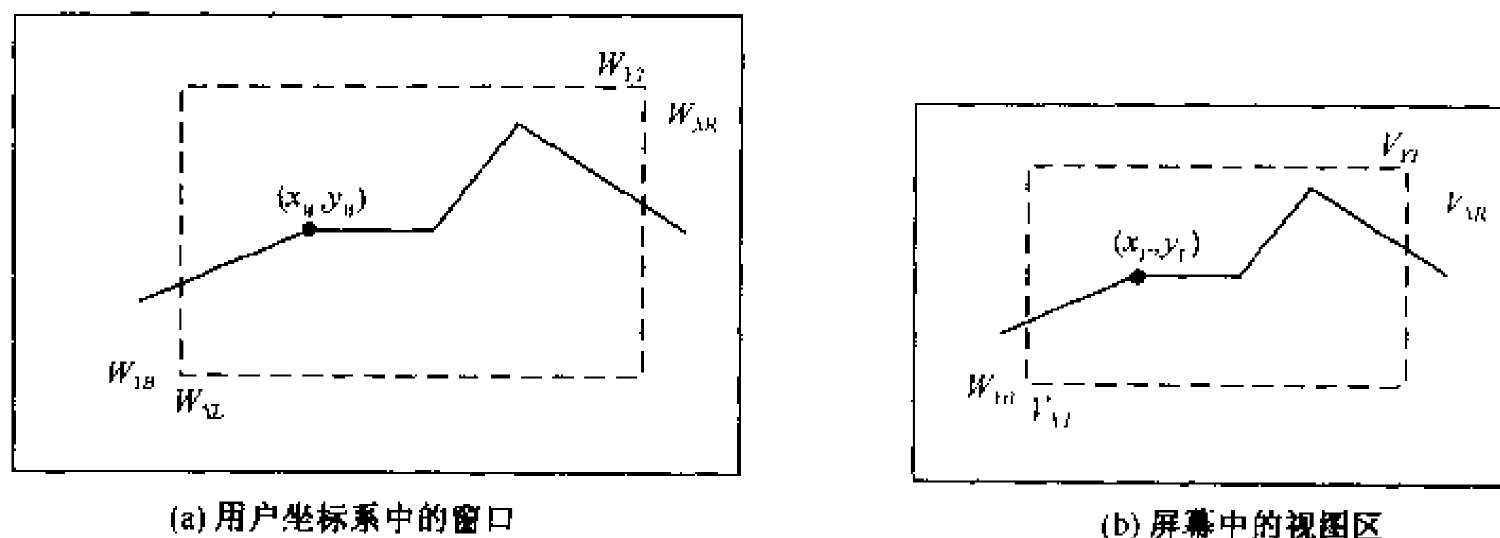


图 9-7 用户坐标系中的窗口与屏幕中视图区的对应关系

$$[x_V y_V 1] = [x_W y_W 1] \times \begin{bmatrix} a & 0 & 0 \\ 0 & c & 0 \\ b & d & 1 \end{bmatrix} \tag{9-14}$$

式中：

$$a = \frac{V_{XR} - V_{XL}}{W_{XR} - W_{XL}};\ b = V_{XL} - W_{XL} \times \frac{V_{XR} - V_{XL}}{W_{XR} - W_{XL}};$$

$$c = \frac{V_{YT} - V_{YB}}{W_{YT} - W_{YB}};\ d = V_{YB} - W_{YB} \times \frac{V_{YT} - V_{YB}}{W_{YT} - W_{YB}}。$$

当 $a \neq c$ 时，即 x 方向与 y 方向的图形变化比例不一致，视图区中的图形会产生变形；当 $a = c = 1$，$b = d = 0$，且窗口与视图区中的坐标原点也相同时，则视图区中将产生与窗口区中相同的图形。

2. 变换过程

二维图形从窗口区到视图区的输出过程如图 9-8 所示，其中 WC、NDC、DC 分别对应用户坐标、规格化坐标和设备坐标。与二维情况类似，三维图形一般需要经过三维裁剪之后，将落在三维窗口内的形体经投影变换，变成二维图形，再在指定的视图区内

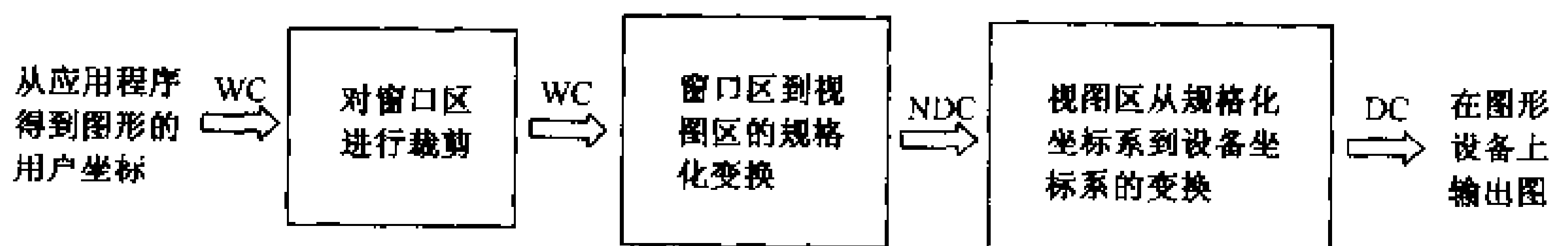


图 9-8 窗口-视图二维变换
（孙家广，1998）

输出，其输出过程如图 9-9 所示，其中 VC 为视窗坐标。

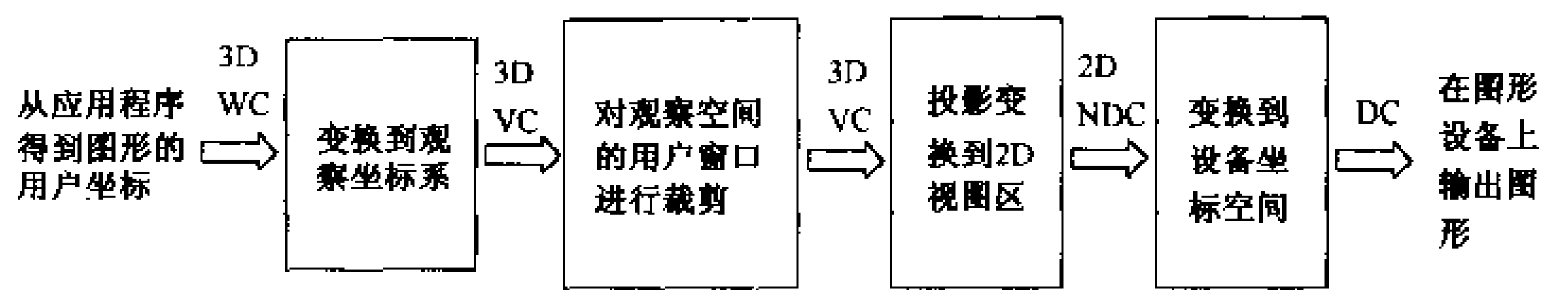


图 9-9 窗口-视图三维变换

（孙家广，1998）

9.4.2 规格化坐标到设备坐标的变换

在窗口-视图的二维变换和三维变换中都需要将规格化坐标变换为设备坐标，即显示器的像素坐标，该变换关系如图 9-10 所示。对于大分辨率为 1024×768（$N_x=1024$，$N_y=768$）的显示器而言，一般地 $a=1$，在 NDC 中的点（x_{NDC}，y_{NDC}）经过平移（d_x，d_y）和比例变换（s_x，s_y）后，就可以得到 DC 中点（x_{DC}，y_{DC}）的坐标如下：

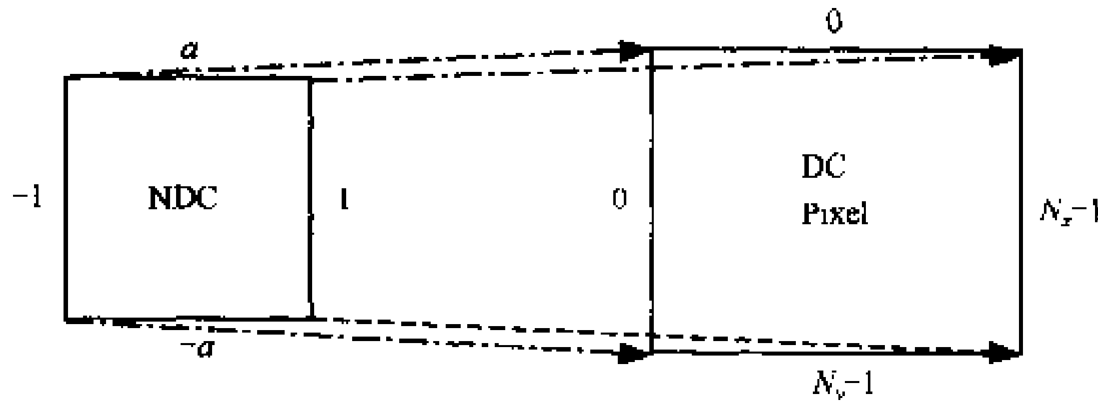


图 9-10 NDC 到 DC 变换的一般关系

（据孙家广，1998）

$$\begin{cases} x_{DC} = s_x \cdot x_{NDC} + d_x \\ y_{DC} = s_y \cdot y_{NDC} + d_y \end{cases} \tag{9-15}$$

值得注意的是，由于实际应用中 NDC 和 DC 的方向是相反的。此时，应结合屏幕的分辨率来计算上式中的两对参数（d_x，d_y）和（s_x，s_y）。

由于在 x 方向上 -1 变成 0，1 变成 N_x-1，而 y 方向上 $-a$ 变成 N_y-1，a 变成 0，因此，两对参数（d_x，d_y）和（s_x，s_y）的计算公式为：

$$\begin{cases} d_x = (N_x - 1)/2; d_y = (N_y - 1)/2 \\ s_x = (N_x - 1)/2; s_y = (N_y - 1)/(-2a) \end{cases} \tag{9-16}$$

9.4.3 二维图形的几何变换

图形变换一般指图形的几何信息经过几何变换后产生新的图形，包括坐标系不动而图形变动和图形不动而坐标系变动两种情况。此处主要讨论图形拓扑关系不发生变化的

几何变换，并由此产生各种形状的复杂图形。

设二维图形变换前任一点坐标的矩阵表示为 $[x\ y\ 1]$，变换后为 $[x'\ y'\ 1]$。

1. 二维变换矩阵

二维图形的几何变换矩阵可用下式表示（孙家广，1998）：

$$T_{2D}=\left[\begin{array}{cc:c} a & d & g \\ b & e & h \\ \hdashline c & f & i \end{array}\right] \tag{9-17}$$

从变换功能上，可以将 T_{2D} 分成 4 个子矩阵，其中：

(1) $\begin{bmatrix} a & d \\ b & e \end{bmatrix}$ 是对图形进行缩放、旋转、对称、错切等变换；

(2) $[c\quad f]$ 是对图形进行平移变换；

(3) $\begin{bmatrix} g \\ h \end{bmatrix}$ 是对图形作投影变换，g 的作用是在 x 轴的 $1/g$ 处产生一个灭点，h 的作用是在 y 轴的 $1/h$ 处产生一个灭点；

(4) $[i]$ 是对整个图形作伸缩变换。

若 T_{2D} 被定义为二维直角坐标系中的单位矩阵，则 T_{2D} 可以被看成 3 个行向量，其中 $[1\ 0\ 0]$ 表示 x 轴上的无穷远点，$[0\ 1\ 0]$ 表示 y 轴上的无穷远点，$[0\ 0\ 1]$ 表示坐标原点。

2. 平移变换

$$[x'\quad y'\quad 1]=[x\quad y\quad 1]\cdot\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_x & d_y & 1 \end{bmatrix}=[x+d_x\quad y+d_y\quad 1] \tag{9-18}$$

3. 比例变换

$$[x'\quad y'\quad 1]=[x\quad y\quad 1]\cdot\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}=[x\cdot s_x\quad y\cdot s_y\quad 1] \tag{9-19}$$

(1) 当 $s_x=s_y=1$，为恒等比例变换，即图形不变；

(2) 当 $s_x=s_y>1$，图形沿两个坐标轴方向等比例放大；

(3) 当 $s_x=s_y<1$，图形沿两个坐标轴方向等比例缩小；

(4) 当 $s_x\neq s_y$，图形沿两个坐标轴方向作非均匀的比例变换。

4. 对称变换

$$[x'\quad y'\quad 1]=[x\quad y\quad 1]\cdot\begin{bmatrix} a & d & 0 \\ b & e & 0 \\ 0 & 0 & 1 \end{bmatrix}=[ax+by\quad dx+ey\quad 1] \tag{9-20}$$

(1) 当 $b=d=0, a=-1, e=1$，有 $x'=-x, y'=y$，产生与 y 轴对称的反射图形；

(2) 当 $b=d=0, a=1, e=-1$, 有 $x'=x, y'=-y$, 产生与 x 轴对称的反射图形；

(3) 当 $b=d=0, a=e=-1$, 有 $x'=-x, y'=-y$, 产生与原点对称的反射图形；

(4) 当 $b=d=1, a=e=0$, 有 $x'=y, y'=x$, 产生与直线 $y=x$ 对称的反射图形；

(5) 当 $b=d=-1, a=e=0$, 有 $x'=-y, y'=-x$, 产生与直线 $y=-x$ 对称反射图形。

5. 旋转变换

$$[x' \quad y' \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= [x \cdot \cos\theta - y\sin\theta \quad x \cdot \sin\theta + y \cdot \cos\theta \quad 1] \tag{9-21a}$$

其中：θ 为 XOY 平面上二维图形绕原点逆时针方向的旋转角度。

若 θ 为 XOY 平面上二维图形绕原点顺时针方向的旋转角度，则上式变为：

$$[x' \quad y' \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= [x \cdot \cos\theta + y\sin\theta \quad -x \cdot \sin\theta + y \cdot \cos\theta \quad 1] \tag{9-21b}$$

6. 错切变换

$$[x' \quad y' \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} 1 & d & 0 \\ b & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [x+by \quad dx+y \quad 1] \tag{9-22}$$

(1) 当 $d=0$, 有 $x'=x+by$, $y'=y$, 此时图形的 y 坐标不变，x 坐标随初值 (x, y) 及变换系数 b 而作线性变化；若 $b>0$, 图形沿 $+x$ 方向作错切位移；若 $b<0$, 图形沿 $-x$ 方向作错切位移；如图 9-11 (a) 所示；

(2) 当 $b=0$, 有 $x'=x$, $y'=dx+y$, 此时图形的 x 坐标不变，y 坐标随初值 (x, y) 及变换系数 d 作线性变化；若 $d>0$, 图形沿 $+y$ 方向作错切位移；若 $d<0$,

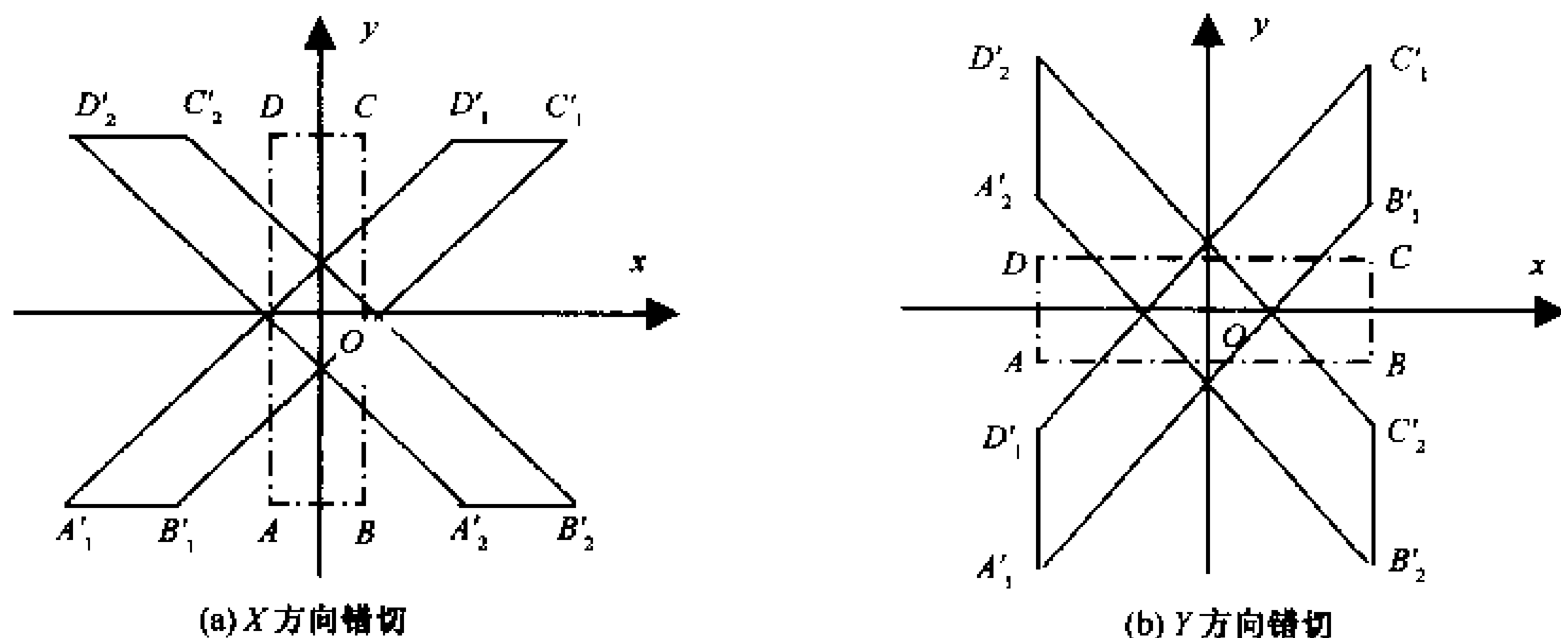


图 9-11 x、y 方向的错切变换
(孙家广，1998)

图形沿 $-y$ 方向作错切位移；如图 9-11（b）所示；

（3）当 $b \neq 0$ 且 $d \neq 0$，有 $x' = x + by$，$y' = dx + y$，图形沿 x、y 两个方向作错切位移。

7. 复合变换

复合变换是指图形作一次以上的几何变换，变换的结果是每次的变换矩阵相乘。

（1）复合平移

$$T_t = T_{t1} \cdot T_{t2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_{x1} & d_{y1} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_{x2} & d_{y2} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_{x1} + d_{x2} & d_{y1} + d_{y2} & 1 \end{bmatrix} \tag{9-23}$$

（2）复合比例

$$T_s = T_{s1} \cdot T_{s2} = \begin{bmatrix} s_{x1} & 0 & 0 \\ 0 & s_{y1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} s_{x2} & 0 & 0 \\ 0 & s_{y2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} s_{x1} \cdot s_{x2} & 0 & 0 \\ 0 & s_{y1} \cdot s_{y2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9-24}$$

（3）复合旋转

$$T_r = T_{r1} \cdot T_{r2} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ -\sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\theta_1 + \theta_2) & \sin(\theta_1 + \theta_2) & 0 \\ -\sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9-25}$$

（4）相对（x_f，y_f）点的比例变换

以上比例、旋转变换均是相对于原点而言的。而实际上，可能需要相对于某一任意点（x_f，y_f）作比例和旋转变换。其变换过程是先把坐标系的原点平移至（x_f，y_f）处，再在新的坐标系下作比例或旋转变换，然后再将坐标原点平移回去。其变换矩阵如下：

$$T_{sf} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -x_f & -y_f & 1 \end{bmatrix} \cdot \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x_f & y_f & 1 \end{bmatrix}$$

$$= \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ (1 - s_x) \cdot x_f & (1 - s_y) \cdot y_f & 1 \end{bmatrix} \tag{9-26}$$

(5) 相对（x_f, y_f）点的旋转变换

$$T_{rf}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -x_f & -y_f & 1\end{bmatrix}\cdot\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ x_f & y_f & 1\end{bmatrix}$$

$$=\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ (1-\cos\theta)\cdot x_f+\sin\theta\cdot y_f & (1-\cos\theta)\cdot y_f-\sin\theta\cdot x_f & 1\end{bmatrix} \qquad (9\text{-}27)$$

9.4.4 三维图形的几何变换

设三维图形变换前任一点坐标的矩阵表示为 [x y z 1]，变换后为 [x' y' z' 1]。

1. 三维变换矩阵

三维图形的几何变换矩阵可用下式表示（孙家广，1998）：

$$T_{3D}=\left[\begin{array}{ccc:c}a_{11} & a_{12} & a_{13} & a_{14}\\ a_{21} & a_{22} & a_{23} & a_{24}\\ a_{31} & a_{32} & a_{33} & a_{34}\\ \hdashline a_{41} & a_{42} & a_{43} & a_{44}\end{array}\right] \qquad (9\text{-}28)$$

从变换功能上，可以将 T_{3D} 分成 4 个子矩阵，其中：

(1) $\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$ 是对图形进行缩放、旋转、错切等几何变换；

(2) $[a_{41}\quad a_{42}\quad a_{43}]$ 是对图形进行平移变换；

(3) $\begin{bmatrix}a_{14}\\ a_{24}\\ a_{34}\end{bmatrix}$ 是对图形作投影变换；

(4) $[a_{44}]$ 是对整个图形产生比例变换。

同理，若 T_{3D} 被定义为三维直角坐标系中的单位矩阵，则 T_{3D} 可以被看成 4 个行向量，其中 [1 0 0 0] 表示 x 轴上的无穷远点，[0 1 0 0] 表示 y 轴上的无穷远点，[0 0 1 0] 表示 z 轴上的无穷远点，[0 0 0 1] 表示坐标原点。

2. 平移变换

$$[x'\quad y'\quad x'\quad 1]=[x\quad y\quad x\quad 1]\cdot\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ d_x & d_y & d_z & 1\end{bmatrix}$$

$$=[x+d_x\quad y+d_y\quad z+d_z\quad 1] \qquad (9\text{-}29)$$

3. 比例变换

若比例变换的参考点为（x_f，y_f，z_f），则与二维坐标变换类似，其比例变换的步骤为：①先把坐标原点平移至参考点；②再在新的坐标系下作比例或旋转变换；③然后再将坐标原点平移回去。其变换矩阵如下：

$$
T_{sf}=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\-x_f&-y_f&-z_f&1\end{bmatrix}\cdot\begin{bmatrix}s_x&0&0&0\\0&s_y&0&0\\0&0&s_z&0\\0&0&0&1\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\x_f&y_f&z_f&1\end{bmatrix}
$$

$$
=\begin{bmatrix}s_x&0&0&0\\0&s_y&0&0\\0&0&s_z&0\\(1-s_x)\cdot x_f&(1-s_y)\cdot y_f&(1-s_z)\cdot z_f&1\end{bmatrix}\tag{9-30}
$$

（1）当 $s_x=s_y=s_z=1$，为恒等比例变换，即图形不变；

（2）当 $s_x=s_y=s_z>1$，图形沿三个坐标轴方向等比例放大；

（3）当 $s_x=s_y=s_z<1$，图形沿三个坐标轴方向等比例缩小；

（4）当 $s_x\neq s_y\neq s_z$，图形沿三个坐标轴方向作非均匀的比例变换。

4. 绕坐标轴的旋转变换

在右手系下，相对于坐标原点绕坐标轴旋转 θ 角的变换公式为：

（1）绕 x 轴旋转

$$
[x'\quad y'\quad z'\quad 1]=[x\quad y\quad z\quad 1]\cdot\begin{bmatrix}1&0&0&0\\0&\cos\theta&\sin\theta&0\\0&-\sin\theta&\cos\theta&0\\0&0&0&1\end{bmatrix}\tag{9-31a}
$$

（2）绕 y 轴旋转

$$
[x'\quad y'\quad z'\quad 1]=[x\quad y\quad z\quad 1]\cdot\begin{bmatrix}\cos\theta&0&\sin\theta&0\\0&1&0&0\\\sin\theta&0&\cos\theta&0\\0&0&0&1\end{bmatrix}\tag{9-31b}
$$

（3）绕 z 轴旋转

$$
[x'\quad y'\quad z'\quad 1]=[x\quad y\quad z\quad 1]\cdot\begin{bmatrix}\cos\theta&\sin\theta&0&0\\-\sin\theta&\cos\theta&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\tag{9-31c}
$$

5. 绕任意轴的旋转变换

如图 9-12 所示，设旋转轴 AB 由空间任一点 A（x_a，y_a，z_a）及其方向数（a，

b，c）定义，空间一点 P（x_p，y_p，z_p）绕 AB 轴旋转 θ 角到 P'（x'_p，y'_p，z'_p）。旋转变换的基本思想为：以 A（x_a，y_a，z_a）为新的坐标原点，并使 AB 分别绕 X 轴、Y 轴旋转适当角度与 Z 轴重合（如图 9-12 所示），再绕 Z 轴转 θ 角，最后再做上述变换的逆变换，使之回到原点的位置。

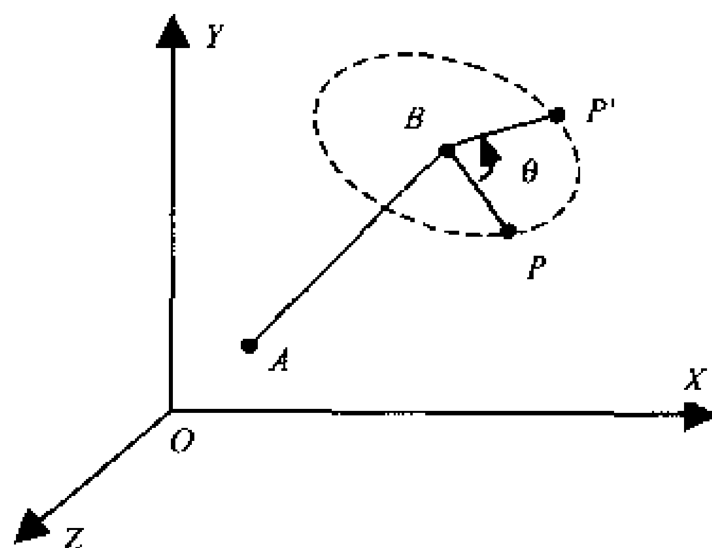


图 9-12　P 点绕 AB 轴旋转

（孙家广，1998）

经过以上步骤，就可以求出旋转变换矩阵 R_{ab}（详见孙家广，1998），进而按下式计算旋转之后 P'（x'_p，y'_p，z'_p）的坐标：

$$[x'_p \quad y'_p \quad z'_p \quad 1] = [x_p \quad y_p \quad z_p \quad 1] \cdot R_{ab} \tag{9-32}$$

参 考 文 献

边馥苓．1996．地理信息系统原理和方法．测绘出版社．63～74

陈俊勇，李京伟．2000．走向 21 世纪的测绘新技术．中国测绘报．2～25

陈述彭等．1999．地理信息系统导论．科学出版社．62～73

龚健雅．2001．地理信息系统基础．科学出版社．159～166

胡　鹏，吴艳兰，杨传勇，李国建．2001．大型 GIS 与数字地球的空间数学基础研究．武汉大学学报·信息科学版．26（4）：296～302

胡毓钜，龚建文，黄　伟．1997．地图投影．测绘出版社

孙家广．1998．计算机图形学．清华大学出版社．362～372

杨启和．1988．关于组合方位投影．测绘通报，（2）：32～34

杨启和．1989．地图投影变换的原理．解放军出版社．

杨晓梅，杨启和，赵　琦．1998．一类新的变比例尺地图投影：组合投影研究．武汉测绘科技大学学报．24（2）：162～165

Peter H. et al., 1997. The Global Spatial Data Infrastructure Initiative and Its Relationship to the Vision of a Digital Earth Toward Digital Earth, 14～20

Snyder J. P., 1987. Magnifying-Glass Azimuthal Map Projection American Cartographer, (1)

下篇 GIS 算法

第十章 空间数据压缩算法

在数据采样过程中，不可避免产生一些数据冗余；大比例尺、高精度数据做小比例尺、低精度应用时，更会存在不必要的数据冗余。为了节约存储空间，如何对采样数据进行合理的删减，以及如何根据问题的需要消除冗余数据，是 GIS 空间数据处理过程中的一项首要任务。

§10.1 基于矢量的压缩算法

GIS 中矢量数据的获取主要有多种途径，包括野外数字化测量、解析测图、数字化仪采集、扫描矢量化等。在解析测图仪的连续方式采样和数字化仪的流方式采样过程中，存在数据密集的问题；在扫描矢量化时，直接由栅格矢量化得到的点也比较密集，可以做合理的删减处理。因此，基于矢量的空间数据压缩的核心是在不扰乱拓扑关系的前提下，对采样点进行合理的删减。

10.1.1 曲线的数据压缩算法

以图 10-1 所示为例，经数据采样得到的曲线 MN 由点序 $\{P_1, P_2, P_3, \cdots, P_n\}$ 组成，这 n 个点的坐标集为 $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_n, y_n)\}$。其中，$P_1$、$P_n$ 分别对应曲线的起点 M 和终点 N。

现根据应用需要和制图比例尺精度要求，给定控制数据压缩的极差为 ε，ε 表示为

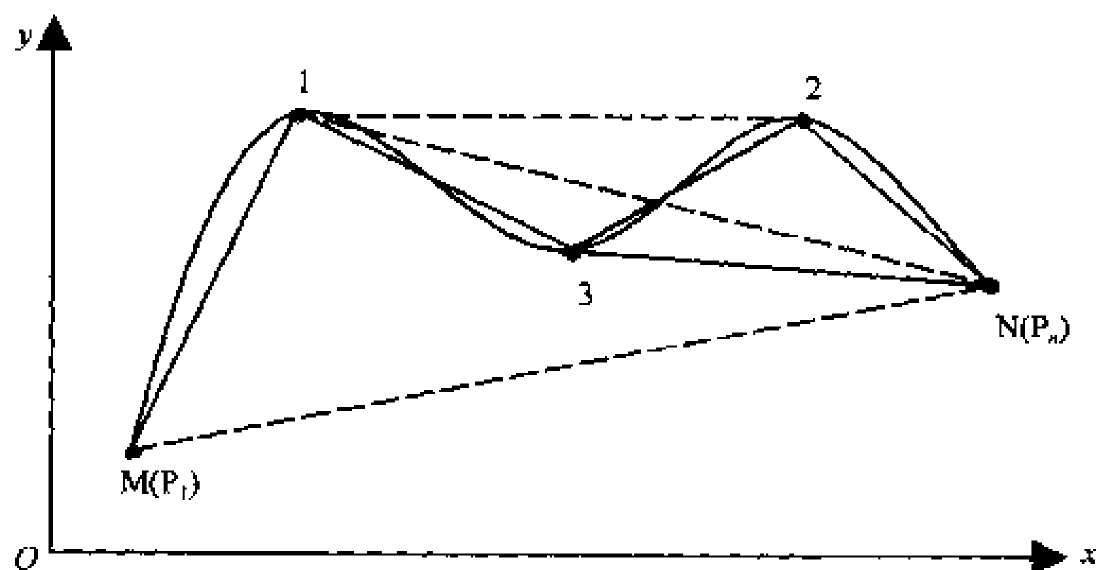


图 10-1 矢量曲线的数据压缩过程实例

被舍弃的点偏离特征点连线之间的垂直距离，一般取为图上的 0.2mm（若比例尺为 1:10000，则实际距离为 2 m）。曲线的空间数据压缩过程如下：

第一步：确定曲线 MN 对应弦的直线方程

根据两点式直线方程，由起点 M、终点 N 建立直线 MN 方程为：

$$\frac{y - y_M}{x - x_M} = \frac{y_M - y_N}{x_M - x_N} \tag{10-1}$$

将上式化简为一般形式为：

$$Ax + By + C = 0 \tag{10-2}$$

其中：

$$A = \frac{y_M - y_N}{\sqrt{(y_M - y_N)^2 + (x_M - x_N)^2}}$$

$$B = \frac{x_M - x_N}{\sqrt{(y_M - y_N)^2 + (x_M - x_N)^2}}$$

$$C = \frac{x_M y_N - x_N y_M}{\sqrt{(y_M - y_N)^2 + (x_M - x_N)^2}}$$

第二步：求曲线 MN 上各点 P_i 到弦线 MN 的距离 d_i

根据点到直线的距离计算公式，P_i（x_i，y_i）到弦线 MN 的距离为：

$$d_i = |Ax_i + By_i + C| \tag{10-3}$$

第三步：求距离 d_i 的最大值 d_h

$$d_h = \max(d_1, d_2, d_3, \cdots, d_n) \tag{10-4}$$

第四步：比较 d_h 与 ε 的大小，并计算开关量 Q

$$Q = \begin{cases} 1 \cdots d_h > \varepsilon \\ 0 \cdots d_h \leqslant \varepsilon \end{cases} \tag{10-5}$$

第五步：决定取舍，提取中间特征点

（1）如果 $Q=0$，则直接可以用弦线 MN（M、N 为特征点）代替曲线 MN；转第六步；

（2）如果 $Q=1$，则将 d_h 所对应的点 P_j（x_j，y_j）抽出，暂时作为中间特征点；然后连接新弦线 MP_j；转第一步（以 MP_j 代替 MN，继续计算和判断）：

①若 $Q=0$，则可以用弦线 MP_j 代替曲线 MP_j；将 P_j 作为中间特征点取出；顺序排在 M 点之后，成为继 M 之后的第一个中间特征点；并连接 P_jN，转第一步（以 P_jN 代替 MN，继续计算和判断）……

②若 $Q=1$，则不可以用弦线 MP_j 代替曲线 MP_j；找到此时 d_h 所对应的点 P_k，并连接新弦线 MP_k；转第一步（以 MP_k 代替 MN，继续计算和判断）……

第六步：形成新的数据文件

将所有提取出的中间特征点从起点 M 开始，顺序排列至终点 N，并写入新的数据文件，即得到化简后的折线的数据文件。

如图 10-1 所示，曲线 MN 的特征点提取过程如下：

首先，找到曲线 MN 上之 d_h 的对应点位为 1 号点；经判断可以用弦线 M1 代替曲线 M1，故 1 号点是继 M 点之后提取出的第一个特征点；

然后，连接弦线 1N；经判断，不可以用弦线 1N 代替曲线 1N；找到曲线 1N 上之 d_h 的对应点位为 2 号点；

故，连接 1、2 号点之弦线 12；经判断，还是不可以用弦线 12 代替曲线 12；找到曲线 12 上之 d_h 的对应点位为 3 号点；

再连接 1、3 号点之弦线 13；经判断，可以用弦线 13 代替曲线 13；故 3 号点是继 1 号点之后提取出的第二个特征点；

进而连接弦线 3N；经判断，不可以用弦线 3N 代替曲线 3N；找到曲线 3N 上之 d_h 的对应点位仍为 2 号点；

然后，连接 3、2 号点之弦线 32；经判断，可以用弦线 32 代替曲线 32；故 2 号点是继 1 号点、3 号点之后提取出的第三个特征点；

再连接 2、N 号点之弦线 2N；经判断，可以用弦线 2N 代替曲线 2N；中间特征点提取结束。

至此可知，曲线 MN 可以用特征点 M、1、3、2、N 顺序连接的折线简化表示。

10.1.2 面域的数据压缩算法

面域空间数据的压缩过程可以看成是组成其边界的曲线段的分别压缩，每段边界曲线的压缩过程同上节所述。但有两个问题需要注意。

1. 封闭曲线的数据压缩

面域由首尾相连的封闭曲线组成。此时，可以人为地将该封闭线分割为首尾相连的两段曲线，然后就可以按前述方法进行压缩。曲线分割的原则是：

（1）原节点是分割点之一；

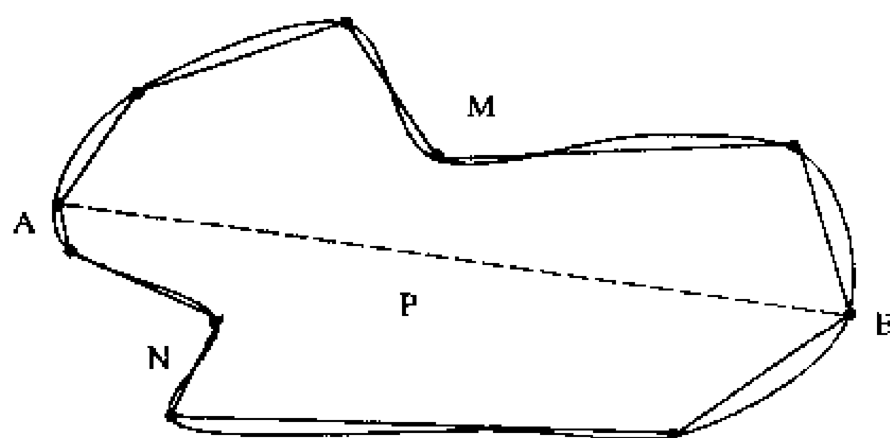


图 10-2 封闭曲线的数据压缩及其结果

(2) 离原节点最远的下一节点是分割点之二。

如图 10-2 所示，多边形 P 的边界曲线由从节点 A 出发的曲线封闭而成；其中曲线上 B 点离节点 A 最远。因而，多边形 P 的边界曲线可以分割为 AMB 和 BNA 两段，进而对曲线段 AMB、BNA 分别进行压缩。

据上述原则，以圆曲线为例进行数据压缩。图 10-3 所示为采用不同的 ε 后的压缩结果。

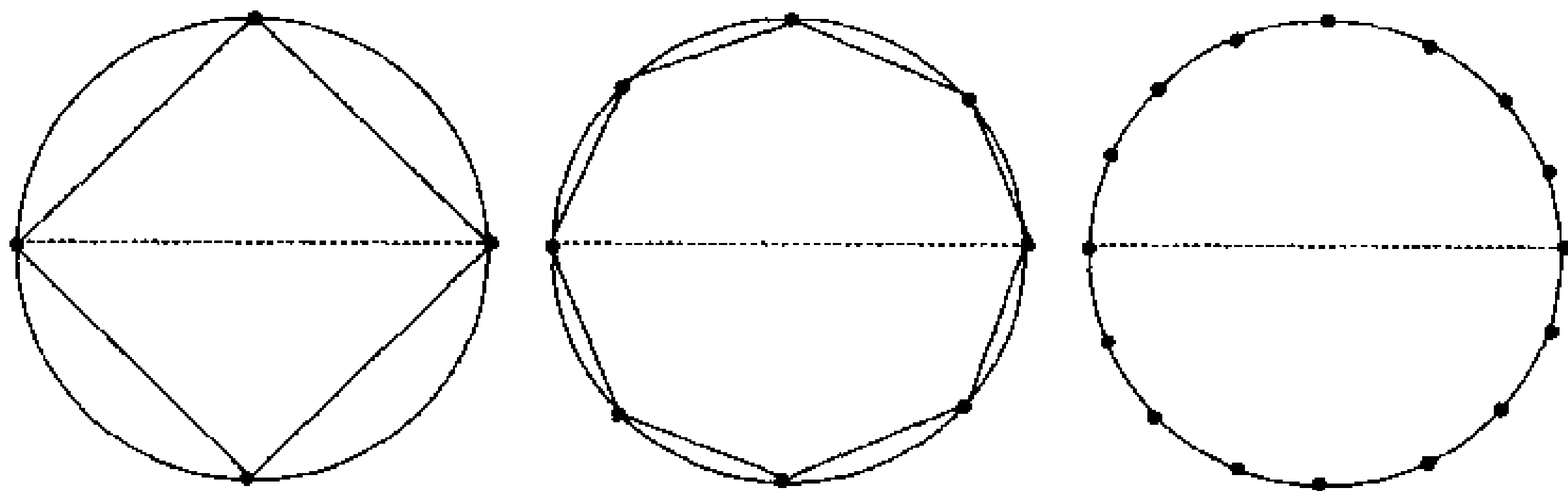


图 10-3 不同 ε 下圆曲线的压缩结果

2. 公共节点的取舍问题

在某些特定情况下，面域的边界曲线由多段首尾相连的曲线连接而成，其压缩可以是分段进行。此时各段曲线的起点和终点必然作为特征点提取出来，因而可能产生数据冗余。比如，当前后曲线段过渡时很平缓，曲线段的公共节点可以不成为特征点，即该点前后的两段曲线可以直接用该点前后的两个特征点的连线来代替。如图 10-4 所示，1、2 号点分别是面域 P 的边界曲线 AB、BC 段的内部特征提取点，因而可以用弦 1B、B2 分别代替曲线 1B 和 B2。而实际上，整个曲线 1B2 仍可以用弦 12 来代替。

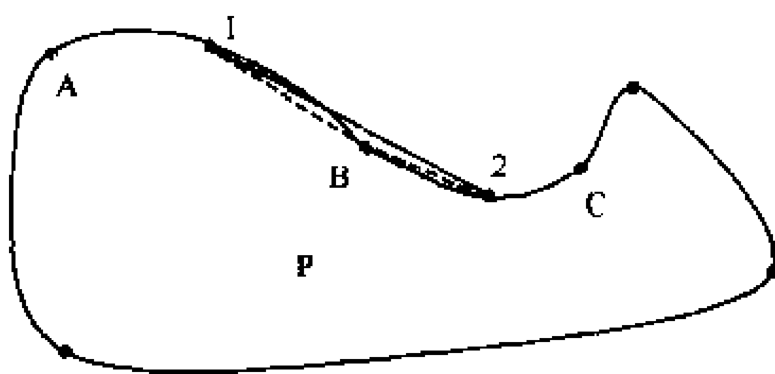


图 10-4 曲线段公共节点的取舍

因此，在处理面域空间数据压缩时，可以在边界曲线分段压缩的基础上，增加一个步骤：即对边界曲线的端点进行可删性检验：

(1) 如果前一曲线最后提取的中间特征点与后一曲线最先提取的中间特征点之间的曲线满足极差控制条件，则两条曲线的连接节点可以删减；

(2) 否则，不可删减。

由于各段边界曲线的数据文件要重新生成，所以当两段曲线的公共节点删减之后，

相当于两条曲线合并为一条曲线。此时可能会扰乱拓扑关系（如曲线AB或BC为多边形的公共边，或AB和BC均为多边形的公共边），因此在处理公共节点的取舍时要慎重，应该对此加以限制。随着计算机能力和网络带宽的提高，只要存储空间和数据传输速率不受限制，一定的数据冗余还是允许的和合理的。

§10.2 基于栅格的压缩算法

如3.2节所述，栅格数据文件记录有3基本方式：基于像元，基于层和基于面域。这3种方式都离不开对像元坐标和属性的记录。因此基于栅格的空间数据压缩的核心是尽量减少像元数量的存储，其方法有三大类，即从减少记录像元的数量入手，或从减少像元的记录信息量入手，以及两者的结合。实用方法有游程长度压缩、差分映射压缩、常规四叉树压缩、线性四叉树压缩和二维行程压缩等。

10.2.1 游程长度压缩算法

假设目标区为M行N列。若基于像元进行记录，则每个像元对应一条记录（R，C，A_1，A_2，…，A_s），其中R为行号，C为列号，A_i为第i层（笛卡尔平面）的属性编码，s为层数。若基于层进行记录，则每个像元对应一条记录（R，C，A_i）。游程长度压缩法就是针对后者的。

可以将同一行中属性相同的若干连续像元视为一个游程，并将其记录为（R，C_r，A_i），其中C_r为该游程最右侧像元的列号。则其中每一行可以简化为K条记录，显然$K \leqslant N$，实现了数据压缩。称之为游程长度压缩法。

以图10-5所示某区域土壤厚度（单位：cm）的栅格数据层为例，若采用未经压缩的基于层的数据记录方式，则需要记录40条记录；若采用游程长度压缩法，只需14条记录，如图10-6所示。因此，仅减少对像元的记录次数，数据压缩率就达到65%$\left(\text{即}\dfrac{(40-14)}{40}\times 100\%\right)$。

120	120	150	150	150	200	200	200
130	130	170	170	170	230	230	230
135	135	135	180	180	180	250	250
140	140	140	200	200	200	200	270
145	145	145	210	210	210	210	210

图10-5 某区域土壤厚度（单位：cm）的栅格数据层

120	0	30	30	30	80	80	80
130	0	40	40	40	100	100	100
135	0	0	45	45	45	115	115
140	0	0	60	60	60	60	130
145	0	0	65	65	65	65	65

图10-6 以行首属性值为参照对图10-5的差分映射结果

表 10-1　游程长度压缩后的数据记录（C_r：列号；A_t：属性值）

行号	第 1 游程		第 2 游程		第 3 游程	
	C_r	A_t	C_r	A_t	C_r	A_t
1	2	120	5	150	8	200
2	2	130	5	170	8	230
3	3	135	6	180	8	250
4	3	140	7	200	8	270
5	3	145	8	210	—	—

10.2.2　差分映射算法

由于属性数据值在计算机中是以二进制方式存储的，数据越小，所占字节数越少。一个字节所能记录的二进制数为 − 127 ~ 127；两个字节所能记录的二进制数为 − 32767 ~ 32767。如果能设法使研究区域内的部分栅格甚至全部栅格的属性值减少，则可以有效地降低栅格数据文件大小。差分映射法就是一种有效的降低栅格数据文件大小的方法。

所谓差分映射法，就是选择某一参照值对有关栅格的属性值进行求差运算，根据差值得到一个新的栅格数据层。参照值的选择有多种方式，即分行选取和全区选取。若分行选取，则可选为该行首列的属性值，也可以选为该行的属性平均值；若全区选取，则可选为首行首列的属性值，也可以选为全区的属性平均值。

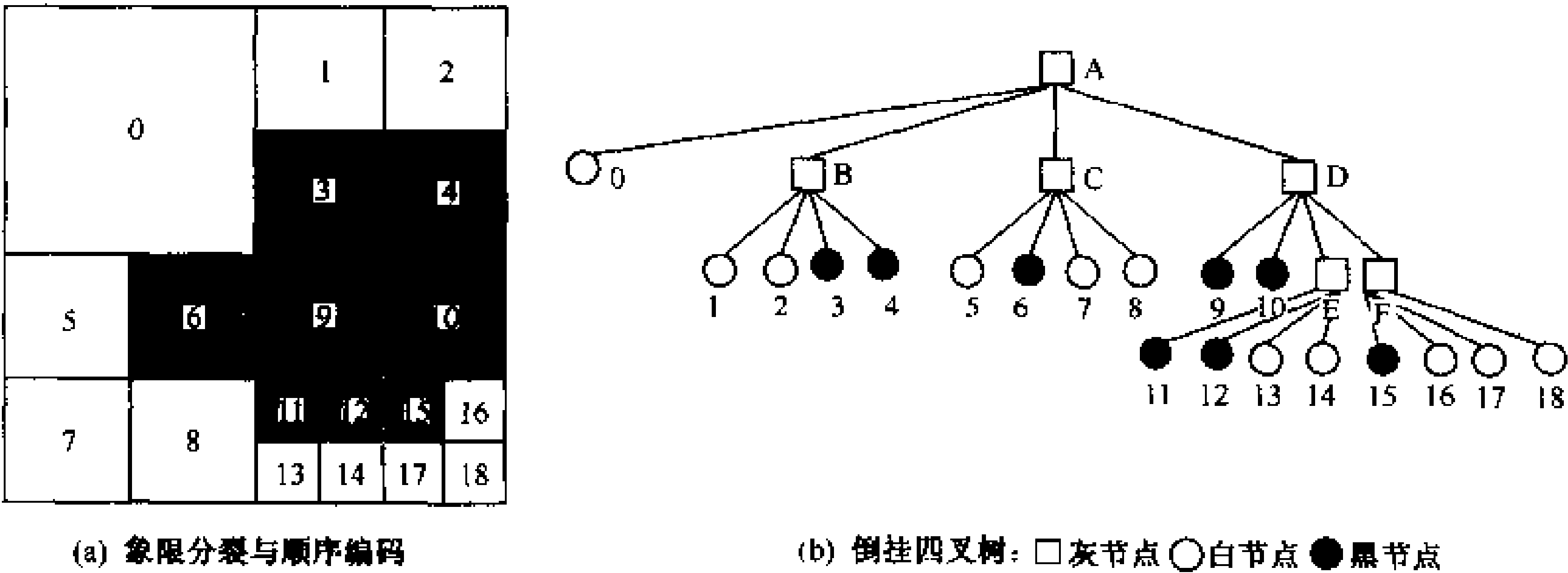
图 10-6 所示为按分行选取方式，以行首属性值为参照，对图 10-5 作差分映射后的结果。可以看出，经差分映射处理后，除第一列外，其余栅格的数据出现为零、位数降低或数字减少。表 10-2 为经差分映射处理前后的各栅格属性记录所需字节数的对比，可见，所需字节数由原来的 79 减少为 44，减少 44.3%。

表 10-2　差分映射前后栅格数据记录长度对比

行号	第 1 列		第 2 列		第 3 列		第 4 列		第 5 列		第 6 列		第 7 列		第 8 列	
	前	后	前	后	前	后	前	后	前	后	前	后	前	后	前	后
1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1
2	2	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1
3	2	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1
4	2	2	2	1	2	1	2	1	2	1	2	1	2	1	2	2
5	2	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1
总	9	9	10	5	10	5	10	5	10	5	10	5	10	5	10	6

10.2.3 常规四叉树压缩算法

常规四叉树（common quadtree, CQT）的基本思想是：首先把一幅图像或栅格地图（$2^k \times 2^k$, $k>1$，不足则补网）等分成四个一级子块，顺序为左上、右上、左下、右下；然后逐块检查其中所有格网属性值（或灰度值），若相同，则该子块不再分；若不同，则将该子块进一步分成四个二级子块；如此递归地分割，直到每个子块的属性或灰度均相等为止。这是由上而下产生四叉树的方法，称为“top-down”法。也可以采用“bottom-up”的方法，即从基本栅格单元开始，按一定的顺序进行检测，如果每相邻四个栅格的属性或灰度值相同，则进行合并，逐次往上递归，直到整幅图均检测完毕。采用以上两种方法建立的四叉树均称为 CQT，如图 10-7 所示。



(a) 象限分裂与顺序编码

(b) 倒挂四叉树：□灰节点 ○白节点 ●黑节点

图 10-7 常规四叉树结构

CQT 的记录信息包括：叶节点（灰节点和黑节点）的属性或灰度值、中间节点（灰节点）及节点之间的联系。每个中间节点需记录 6 个量：1 个指向父节点的指针，4 个指向子节点的指针，本节点的属性值（或灰度值）；每个叶节点需要记录 3 个量：位置、指向父节点的指针及本节点的属性值（或灰度值）。在图 10-7 中，经 3 级分割，共产生 19 个叶节点、5 个中间节点，因此需要 24 条记录，共需 87 个字节。若用基于层的栅格存储方式，则需对 64 个基本栅格分别存储，需要 64 条记录，若每条记录长为 3 个字节（对应 3 个量：行号、列号和属性值），共需 192 个字节。因此，与基于层的栅格存储方式相比，采用 CQT 存储后，所需字节单位由 192 减少为 87，压缩率为 54.7%。

尽管如此，由于频繁地存储和操作指针，影响了压缩效率，增加了操作的复杂性。所以，CQT 一般在数据索引和图幅索引方面得到应用，在数据压缩方面人们多采用线形四叉树（linear quadtree, LQT）方法（Mark D. et al., 1985; Ibbs T., 1988）。

10.2.4 线性四叉树压缩算法

线性四叉树（LQT）只存储最后叶节点（灰节点和黑节点）的信息，包括 3 个量：

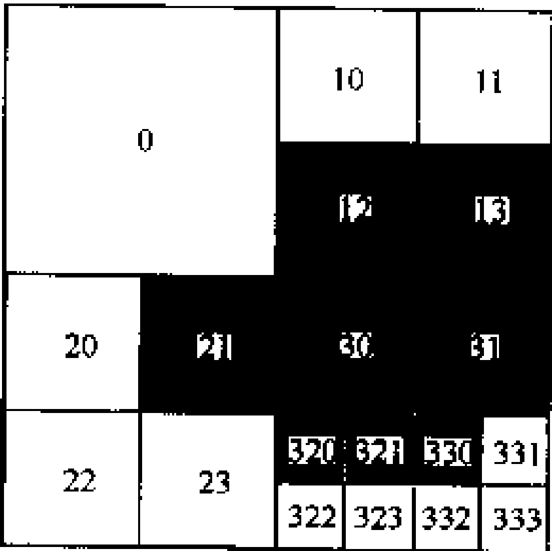
叶节点的位置、大小和属性（或灰度）值。LQT 叶节点的编号称为地址码，它隐含了叶节点的位置信息。最常用的地址码是基于四进制和基于十进制的 Morton 码（Samet，1981；Mark et al.，1985）。

1. 基于四进制的 Morton 码

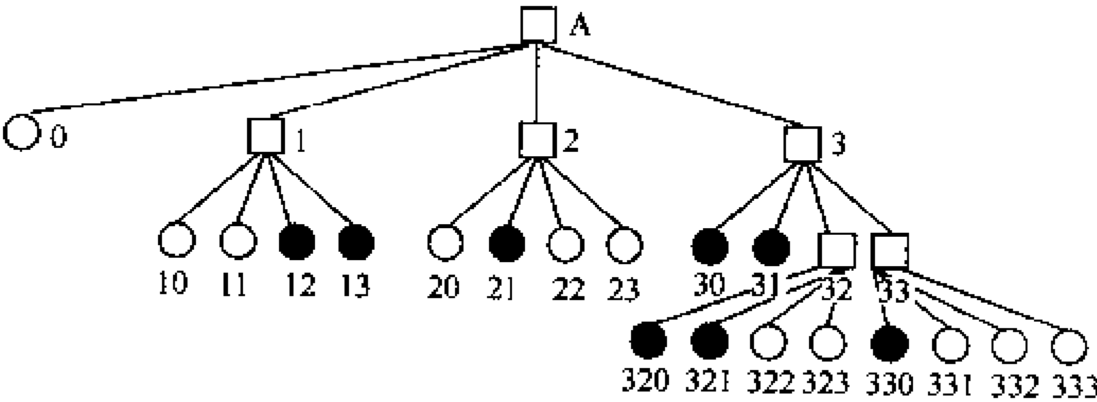
基于四进制的 Morton 码是用 0、1、2、3 共 4 个数字来表示每次分裂后产生左上、右上、左下、右下四象限的标号。在逐级分割过程中，标号的位数不断增加，即每分割一次，标号就增加一位。分割的次数越多，所得到的子区就越小，标号的位数也就越长。所得到的标号，即为任意节点的基于四进制表示的 Morton 码，表示为 M_Q。如图 10-8 所示。最后节点的 Morton 码 M_Q 是所有各位上相应象限值的顺序排列：

$$M_Q = q_1q_2q_3\cdots q_k = q_1 \cdot 10^k + q_2 \cdot 10^{k-1} + q_3 \cdot 10^{k-2} + \cdots + q_k \qquad (10\text{-}6)$$

M_Q 的生成和四叉树的建立有两种不同方案：一是在自上而下（top-down）分裂产生四叉树的过程中自动产生 Morton 码；二是先计算每个格网的 Morton 码，然后按一定的扫描方式自下而上（bottom-up）合并建立四叉树。图 10-8 所示为自上而下分裂产生四叉树的过程和结果。



(a) 象限分割与四进制Morton编码



(b) 倒挂四叉树：□ 灰节点 ○ 白节点 ● 黑节点

图 10-8 基于四进制 Morton 码的线形四叉树结构

由图可见，采用线性四叉树分裂后产生了 19 个叶节点，共需 19 条记录；若每条记录的长度为 3 个字节（对应 3 个量），共需 57 个字节。因此，与基于层的栅格存储方式相比，采用线形四叉树存储后，所需字节单位由 192 减少为 57，压缩率为 70%。

显然，线形四叉树的压缩存储效率是高的。但是，在产生线形四叉树和编码过程中，需要重复检测栅格，每一叶节点的重复检测的次数等于其 M_Q 码的位数，因而运算速度较慢。当栅格区域越大，栅格尺寸越小，区内地物要素越复杂，用这种方法建立线形四叉树的过程越慢。例如，有一个 $2^n \times 2^n$ 的栅格区域，假设其叶节点的 M_Q 平均长度为（$n-2$），则检测格网的次数为 S_{td}：

$$S_{td} = 2 \times 10^{n-2} \qquad (10\text{-}7)$$

因此，人们提出采用自下而上的合并方法来改进运算速度。该法的实质是先将二维

栅格矩阵的行列号（从1开始计数）转换为一维的 M_Q 码，并按升序排列成线性表。将二维栅格矩阵的行列号转换为一维的 M_Q 码的算法为：

$$\left.\begin{aligned} R_b &= \sum_{k=0}^{\log_2(R_d)} \mathrm{MOD}(I_k, 2) \cdot 4^k \\ I_k &= R_d \quad k = 0 \\ I_k &= \mathrm{INT}(I_{k-1}/2) \quad k > 0 \end{aligned}\right\} \tag{10-8}$$

$$\left.\begin{aligned} C_b &= \sum_{k=0}^{\log_2(C_d)} \mathrm{MOD}(I_k, 2) \cdot 4^k \\ I_k &= C_d \quad k = 0 \\ I_k &= \mathrm{INT}(I_{k-1}/2) \quad k > 0 \end{aligned}\right\} \tag{10-9}$$

$$M_Q = 2 \cdot R_b + C_b \tag{10-10}$$

式中：MOD 为取余函数；INT 为取整函数；R_d 为十进制行号；C_d 为十进制列号；R_b 为二进制行号；C_b 为二进制列号；I_k、k 为中间循环变量。

以一个 $2^3 \times 2^3$ 的栅格区域为例，其十进制行列号、二进制行列号和四进制 M_Q 码的对比见表10-3。

表10-3 基于四进制的 Morton 码线性表

M_Q 码 \ 列号 / 行号		C_d	0	1	2	3	4	5	6	7
		C_b	0	1	10	11	100	101	110	111
R_d	R_b									
0	0		000	001	010	011	100	101	110	111
1	1		002	003	012	013	102	103	112	113
2	10		020	021	030	031	120	121	130	131
3	11		022	023	032	033	122	123	131	133
4	100		200	201	210	211	300	301	310	311
5	101		202	203	212	213	302	303	312	313
6	110		220	221	230	231	320	321	330	331
7	111		222	223	232	233	322	323	332	333

在排好的线性表中，从基本栅格开始，依次检查四个相邻的 M_Q 码对应栅格的属性（灰度）值，如果相同则合并为一个大块；否则将其存盘，记录内容包括：M_Q 码，深度和属性（灰度）值。第一轮检测完毕之后，依次检测每四个大块的栅格的属性（灰度）值，若相同则合并；否则将其存盘；继续检测……直到没有可以合并的子块为止。

可以看出，自下而上的合并方法仅对少数大块检测两次或多次，而对大部分基本栅格仅检测一次。对于一个 $2^n \times 2^n$ 的栅格区域，总的检查次数至多为 S_{bu}：

$$S_{bu} = 2^n \times 2^n + 2^{n-1} \times 2^{n-1} + 2^{n-2} \times 2^{n-2} + \cdots 1 = \frac{1}{3} \times 2^{2n+2} \tag{10-11}$$

若取 $n = 10$，由式（10-7）、（10-11）得 $S_{td} = 2 \times 10^8$，$S_{bu} = \frac{1}{3} \times 2^{22}$。可知，后者

是前者的千分之七。

龚健雅（2001）曾利用上述两种方法，分别对3幅不同大小的遥感图像建立线性四叉树，并对建树效率进行了对比，结果如表10-4所示。

表10-4　两种建立线性四叉树方法的效率比较

（龚健雅，2001）

时间 图像 方法	Ⅰ (64×64)	Ⅱ (128×128)	Ⅲ (256×256)
自下而上法	1.20 s	4.23 s	26.08 s
自上而下法	1.42 s	8.42 s	2 min 30.20 s

2. 基于十进制的Morton码

由于基于四进制的Morton码及线性四叉树的建立方法需要开销大量的内外存，采用自下而上法的事先排序会花费大量时间；而且大多数高级语言并不支持四进制变量（虽然可以用十进制长整形来表示Morton码，但很浪费）。因此，基于四进制的Morton码及线形四叉树的建立方法并不实用。

于是，Mark等人（1989）建议采用基于十进制的Morton码（简称M_D码）作为线性四叉树的地址码，并使用自下而上的方法建立四叉树。这样，合并过程直接可以按自然数的顺序进行扫描，可以省去排序的过程；还可以不用记录地址码和深度值，既节约了存储空间，又提高了运算效率。

根据栅格的行列号计算栅格的M_D码有两种方法：数学计算和按位操作计算。

1）M_D码的数学计算

M_D码的数学计算与M_Q码类似。

$$\left.\begin{aligned} R_f &= \sum_{k=0}^{\log_2(R_d)} \mathrm{MOD}(I_k,2)\cdot 10^k \\ I_k &= R_d \quad k=0 \\ I_k &= \mathrm{INT}(I_{k-1}/2) \quad k>0 \end{aligned}\right\} \tag{10-12}$$

$$\left.\begin{aligned} C_f &= \sum_{k=0}^{\log_2(C_d)} \mathrm{MOD}(I_k,2)\cdot 10^k \\ I_k &= C_d \quad k=0 \\ I_k &= \mathrm{INT}(I_{k-1}/2) \quad k>0 \end{aligned}\right\} \tag{10-13}$$

$$M_D = 2\cdot R_f + C_f \tag{10-14}$$

式中：R_f为行号伪码；C_f为列号伪码；其他符号意义同前。

以一个$2^3\times 2^3$的栅格区域为例，其十进制行列号、行列伪码和M_D码的对比见表10-5。

表 10-5 基于十进制的 Morton 码表

M_D码 \ 列号 / 行号		C_d	0	1	2	3	4	5	6	7
		C_f	0	1	4	5	16	17	20	21
R_d	R_f									
0	0		0	1	4	5	16	17	20	21
1	1		2	3	6	7	18	19	22	23
2	4		8	9	12	13	24	25	28	29
3	5		10	11	14	15	26	27	30	31
4	16		32	33	36	37	48	49	52	53
5	17		34	35	38	39	50	51	54	55
6	20		40	41	44	45	56	57	60	61
7	21		42	43	46	47	58	59	62	63

2) M_D 码的按位操作计算

设十进制的行列号在计算机内部的二进制表示分别为：

$$\left.\begin{aligned} R_b &= (r_n r_{n-1} \cdots r_3 r_2 r_1) \\ C_b &= (c_n c_{n-1} \cdots c_3 c_2 c_1) \end{aligned}\right\} \tag{10-15}$$

事实上 R_b、C_b 的二进制码并非等长。此时短者要向长者看齐，即在短者的高位补零。然后，将二进制表示的行列号交叉放置，即得到二进制表示的 Morton 码 M_b 如下：

$$R_b=(r_n r_{n-1} \cdots r_3 r_2 r_1) \qquad C_b=(c_n c_{n-1} \cdots c_3 c_2 c_1)$$
$$M_b=(r_n c_n r_{n-1} c_{n-1} \cdots r_2 c_2 r_1 c_1) \tag{10-16}$$

若采用 C、汇编等高级语言的按位操作模式，则首先依次取出行列号的二进制数 R_b、C_b 中 i（$i=1\sim n$）位上的数字 r_i、c_i 如下：

$$\left.\begin{aligned} r_i &= (R_b \ \& \ 2^{i-1}) \\ c_i &= (C_b \ \& \ 2^{i-1}) \end{aligned}\right\} \tag{10-17}$$

式中：& 为按位操作的“与”运算符。

然后，利用按位操作的“或”运算“|”，将所取出的 r_i、c_i 依次交叉放入 M_b 的变量中，即得到如式（10-16）表示的二进制 Morton 码 M_b。通过二进制转换为十进制，即可以方便地得到十进制表示的 Morton 码 M_D。

图 10-9 所示为基于十进制 Morton 码 M_D 的栅格图像分裂和编码实例。其中合并后子块的 M_D 取为该子块的起始 M_D，即该子块中所有栅格的最小 M_D。对比图 10-8 和图 10-9 可以看出，两者的象限分裂结果及倒挂四叉树形状是完全一样的，只不过后者的子区编号为十进制，而前者为四进制。前已述及，十进制地址的存储要优于四进制。因此，基于十进制 Morton 码 M_D 的 LQT 压缩率应高于四进制 Morton 码 M_Q 的 LQT 压缩率。图 10-9 所示，相对于基于层的栅格存储方式，其压缩率应大于 70%。

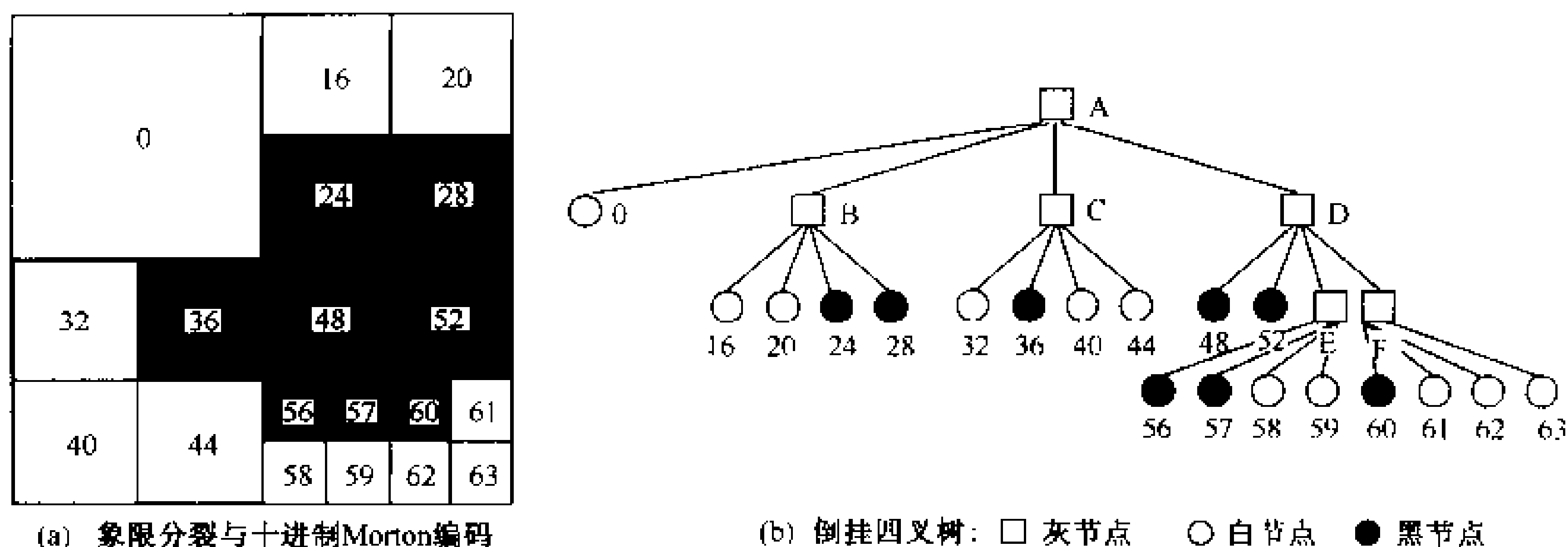


图 10-9 基于十进制 Morton 码的线性四叉树结构

10.2.5 二维行程压缩算法

传统的一维行程（one dimensional run-length, 1DRL）压缩法通过记录入口位置和属性值，依次扫描后续位置的属性值。若后一位置的属性值发生变化（即不等于前一位置的属性值），则记录后一位置的地址和属性值；若后一位置的属性值不发生变化，则不作记录……直到最后，只记录最后位置的地址和属性值即可。这种压缩法避免了对一维连续的属性数据的重复记录，达到了数据压缩的目的。

二维行程（two dimensional run-length, 2DRL）压缩法是在对 1DRL 压缩法的二维扩展的基础上，融合了 LQT 及其 Morton 码技术。它通过对已经形成的基于 M_D 码的栅格图像的 LQT 记录进行扫描，记录其入口 M_D 地址和属性值；然后依次扫描后续地址的属性值。若后一地址的属性值发生变化（即不等于前一地址的属性值），则记录后一地址和属性值；若后一位置的属性值不发生变化，则不作记录……直到最后，并记录最后位置的地址和属性值即可。

图 10-9 所示的 LQT 的栅格数据记录见表 10-6 的左侧，其 2DRL 记录见表 10-6 的右侧。

由表可见，在 2DRL 压缩法中，前后两个地址之差表达了该行程的长度，即属性不变的栅格的长度，亦即表示该子块（不一定是 $2^n \times 2^n$）的大小。显然，2DRL 压缩后的数据已失去了四叉树的概念，但它比规则的四叉树更节省存储空间，而且有利于后续的数据修改，包括插入、删除等。

如图 10-10 所示的图像相当于图 10-9 所示的栅格图像发生了局部变化，变化后的 LQT 的栅格数据记录和 2DRL 记录如表 10-7 所示。可见，表 10-7 是直接对表 10-6 进行简单的插入和修改得到的（表中斜体所示为插入和修改的记录）。

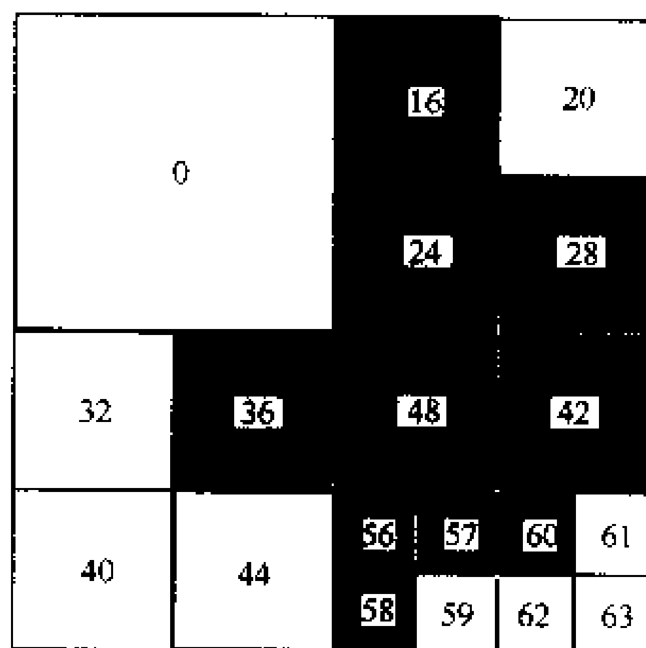


图 10-10 图 10-9 局部变化后的结果

表 10-6　基于 M_D 的 LQT 记录与 2DRL 编码记录的对比

基于 M_D 的 LQT 记录		2DRL 编码记录	
M_D	灰度值	M_D	灰度值
0	0	0	0
16	0	24	1
20	0	32	0
24	1	36	1
28	1	40	0
32	0	48	1
36	1	58	0
40	0	60	1
44	0	61	0
48	1	63	0
52	1		
56	1		
57	1		
58	0		
59	0		
60	1		
61	0		
62	0		
63	0		

表 10-7　局部变化后数据记录插入和修改的结果

基于 M_D 的 LQT 记录		2DRL 编码记录	
M_D	灰度值	M_D	灰度值
0	0	0	0
16	*1*	16	*1*
20	0	*20*	*0*
24	1	24	1
28	1	32	0
32	0	36	1
36	1	40	0
40	0	48	1
44	0	*59*	*0*
48	1	60	1
52	1	61	0
56	1	63	0
57	1		
58	*1*		
59	0		
60	1		
61	0		
62	0		
63	0		

参 考 文 献

龚健雅. 2001. 地理信息系统基础. 北京：科学出版社. 102～112, 131～132

李树祥. 1987. Q码与图像表达. 国防科技大学学报. No.2

杨　凯等. 1988. 遥感图像处理和方法. 北京：测绘出版社

Ibbs T. and Stevens A., 1988. Quadtree storage of Vector Data, Int. J. GIS, No.1

Mark D. and Abel D., 1985. Linear Quadtree from Vector Presentation of Polygons, IEEE, Transactions on Pattern Analysis and Intelligence, No.3

Samet H., 1981. An Algorithm for Converting Raster to Quadtrees. IEEE, Transactions on Pattern Analysis and Intelligence, No.1

Shaffer C. et al., 1990. QUILT: A GIS based on Quadtrees, Int. J GIS, No.2

第十一章　空间数据内插算法

空间数据内插是 GIS 数据处理的另一项重要任务。所谓空间数据内插，就是根据一组已知的离散数据或分区数据，按照某种数学关系推求出其他未知点或未知区域的数据的数学过程。GIS 在很多情况下，必须进行空间数据内插，比如采样密度不够、采样分布不合理、采样存在空白区、等值线的自动绘制、数字高程模型的建立、区域边界分析、曲线光滑处理、空间趋势预测、采样结果的 2.5 D 可视化等。进行空间数据内插的方法多种多样，可以从内插时使用已知采样点的范围分为两大类：整体拟合和局部拟合；也可以从内插的具体内容分为两大类：点的内插和区域内插。

所谓整体拟合，是指内插模型是基于研究区域内的所有采样点的特征观测值建立的，如趋势面分析、傅立叶级数等。整体拟合的特点是不能提供内插区域的局部特性，如金矿品位富集、辐射源等局部异常。所以整体拟合通常用于大范围、长周期变化情况，如沙漠地貌、平原地貌、地下水位、煤层分布、海水间温层、大气对流层等，内插结果一般具有粗略性特点。

所谓局部拟合，是指仅用邻近于未知点的少数已知采样点的特征值来估算该未知点的特征值，如样条函数法、移动平均法等。局部拟合的特点是可以提供内插区域的局部特性，且不受其他区域的内插影响。所以局部拟合通常用于如地下溶洞推测、金属矿品位估计、陷落柱预测、污染源搜索等，内插结果一般具有精确性特点。

以下将围绕点的内插、区域内插和曲线的光滑处理来介绍整体拟合与局部拟合所涉及的算法。

§11.1　点的内插算法

根据整体拟合和局部拟合分类，点的内插算法如表 11-1 所示。以下将介绍其中的主要算法。

表 11-1　点的内插算法

整体拟合法	局部拟合法
趋势面法	Kriging 法
最小二乘法	移动内插法
傅立叶级数法	移动平均法
最小二乘样条函数法	样条函数法
距离加权最小二乘法	双样条函数法
	多面函数法
	线性内插法
	双线性多项式内插法
	有限差分法

11.1.1 基于整体拟合的算法

1. 趋势面法

所谓趋势面法，是通过选择一个二元函数来逼近采样数据的整体变化趋势。该二元函数的一般形式为：

$$f(x, y) = \sum_{r+s=0}^{r+s=p} b_{rs} x^r y^s \tag{11-1}$$

式中：p 为二元函数的阶数。

（1）当 $p=0$ 时，为水平面：

$$f(x, y) = b_0 \tag{11-2}$$

（2）当 $p=1$ 时，为倾斜平面：

$$f(x, y) = b_0 + b_1 x + b_2 y \tag{11-3}$$

可用于模拟边坡、倾斜煤层、断层等；

（3）当 $p=2$ 时，为二次曲面：

$$f(x, y) = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2 \tag{11-4}$$

可用于模拟地形起伏、褶曲煤层等。

该二元函数必须满足观测值与拟合值之差的平方和最小：

$$\sum_{i=1}^{i=n} (z(x_i, y_i) - f(x_i, y_i))^2 = \min \tag{11-5}$$

式中：z（x_i，y_i）为（x_i，y_i）点处的采样值；n 为总采样数。

可以用多重回归技术确定上述各式的系数 b_i。用二元函数进行趋势面内插，具有以下特点：

（1）当 $n>3$ 时，拟合曲面常产生异常大或异常小的值；

（2）拟合残差属正常分布的独立误差，具有一定的相关性；

（3）用于局部内插之前，要事先对宏观异常的采样值进行处理。

2. 最小二乘法

最小二乘法是测量学科中的一种广泛应用的测量平差和数据内插方法。一个测量结果通常包含三部分：①可以用某个函数表示的部分值，这个函数在空间为一个曲面，称为趋势面；②由于仪器等固定问题引起的部分值，它不能简单地用某个函数表示，称为系统误差；③由于偶然现象引起的部分值，称为偶然误差或随机噪音。

以数字地面模型为例，如图 11-1 所示，将某一区域内数据点的高程值 Z 用某一趋势面 z（x，y）拟合后，各点上的余差 D 包含两部分：系统误差 S 和偶然误差 R。这些数据之间具有以下关系：

$$\left.\begin{aligned} Z &= z + D \\ D &= S + R \end{aligned}\right\} \tag{11-6}$$

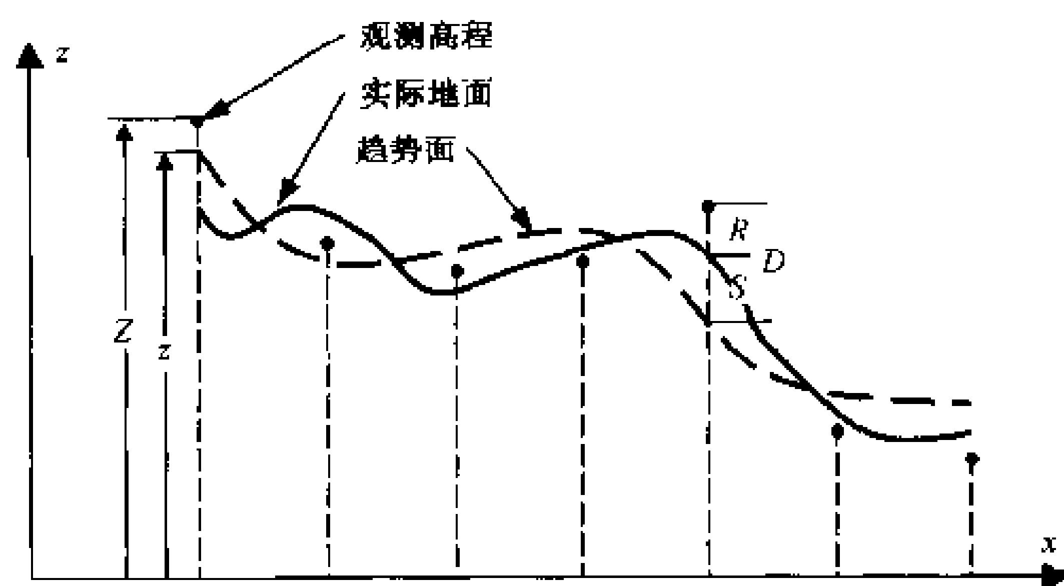


图 11-1　地面高程趋势面与余差

且应满足：

$$\sum(D) = \sum(R)$$

若一个子区内共有 n 个采样点，可以用一个二次多项式曲面来拟合地形，则观测值方程的矩阵表达为：

$$Z = BX + S + R \tag{11-7}$$

式中：Z 为观测值列向量；B 为二次曲面的系数矩阵；X 为二次曲面的参数列向量；S 为系统误差列向量；R 为偶然误差列向量。

若引入 m 个内插点的系统误差 S'，则上式变为：

$$Z = BX + S + OS' + R \tag{11-8}$$

式中：O 为 $n \times m$ 阶的零向量。

对（11-7）和（11-8）式的求解对应三种不同方式：

(1) 推估法（也称预测）：认为内插点的信号 S' 与采样点的信号 S 线性相关。可以采用线性内插的方法，根据 n 个采样点，由式（11-7）确定 n 个点的信号 S，进而内插出待定点的信号；

(2) 滤波法：采用间接平差的方法，根据 n 个采样点，通过滤波滤掉采样点上的噪音，进而确定式（11-7）及 n 个点的信号值 S，最后求出待定点的信号；

(3) 配置法：采用广义平差的方法，对（11-8）式同时解算 n 个采样点的信号 S 和 n 个待定点的信号 S'。

11.1.2　基于局部拟合的算法

1. Kriging 法

Kriging 法又称空间自协方差最佳内插法。该法是南非矿业工程师 D. G. Krige，针对矿业特色与地质推估问题创立的矿石品位最佳内插法。Kriging 法的基本原理与最小二乘法类似。其实质是既考虑采样表面的总趋势变化（统计特征），又考虑采样表面特征的相关变化和随机变化。Kriging 将上述三项分别称为采样表面的结构项、相关项和随机噪音。如图 11-2 所示，其一般拟合函数式为：

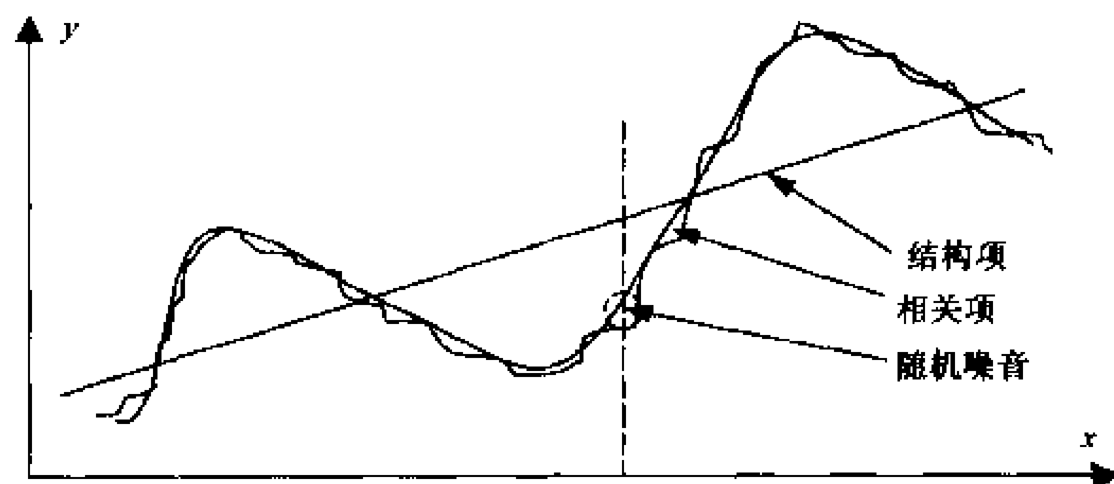


图 11-2 Kriging 法的基本原理

$$f(x,y) = f_1(x,y) + f_2(x,y) + C \tag{11-9}$$

式中：f_1（x，y）为结构项；f_2（x，y）为相关项；C 为随机噪音。

近年来，国内外许多学者对最初的 Kriging 法进行了演绎和推广，已经成为一种系列 Kriging 内插方法的代名词，并且在地质、矿业、地理、水文、土地等领域获得广泛应用。其中普通 Kriging 法与最小二乘的推估法类似，而广泛 Kriging 法与最小二乘的配置法类似。所不同的是两者所采用的协方差函数不同；而且在配置法与推估法中，插值的方差－协方差是严格数学意义上的方差-协方差阵，而 Kriging 法是采用半方差(或称半变异函数)，来构造推估用的方差协方差阵。限于篇幅，具体内容可参看有关文献。

2. 移动内插法

移动内插法是一种典型的逐点内插法。其实质是：首先以内插点 P 为中心，按某一半径 R 作圆；然后选定某一多项式内插函数，用落在该圆内的采样点的特征观测值来拟合该范围的特征值曲面；进而求得待插点的特征值，如图 11-3 所示。多项式内插函数的典型代表为二次多项式：

$$f(x,y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 \tag{11-10}$$

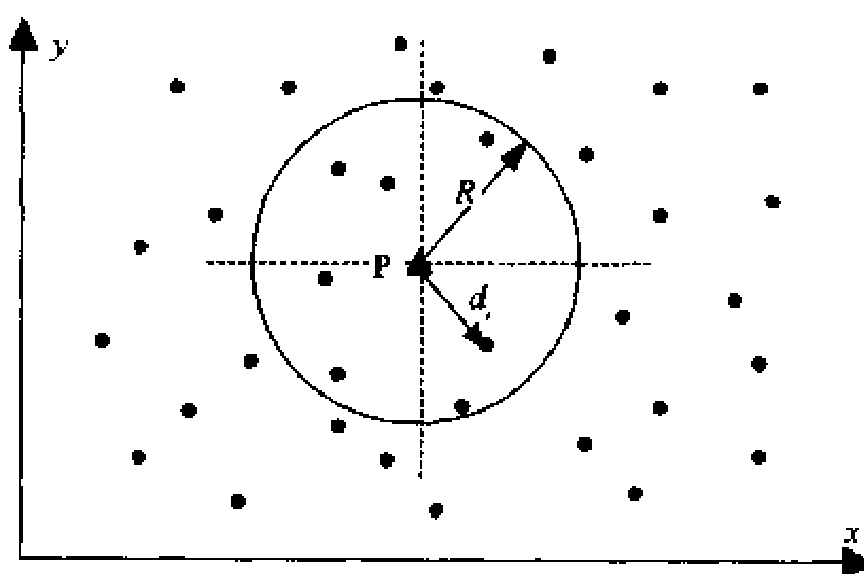


图 11-3 移动内插法取样

上式有六个待定系数，因此只要取样半径内有 6 个采样点（最好是四个象限内均有点），即可以确定这 6 个未知数。当采样点不足 6 个时，需要扩大取样半径；当采样点超过 6 个时，要列出 n 个采样点的误差（v_i）方程的矩阵如下：

$$v = MB - Z \tag{11-11}$$

式中：

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix};\quad M = \begin{bmatrix} 1 & \overline{x_1} & \overline{y_1} & \overline{x_1^2} & \overline{x_1}\,\overline{y_1} & \overline{y_1^2} \\ 1 & \overline{x_2} & \overline{y_2} & \overline{x_2^2} & \overline{x_2}\,\overline{y_2} & \overline{y_2^2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \overline{x_n} & \overline{y_n} & \overline{x_n^2} & \overline{x_n}\,\overline{y_n} & \overline{y_n^2} \end{bmatrix};B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_5 \end{bmatrix};Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

根据平差理论，二次曲面系数的解为：

$$B = (M^TPM)^{-1}M^TPZ \tag{11-12}$$

式中 P 为权重矩阵。考虑到采样点离内插点 P 的距离不同而相关程度不同，可以采用不同的权重 p_i，n 个采样点即构成 $1 \times n$ 的矩阵。p_i 的取定有以下 3 种基本方式：

$$p_i = \frac{1}{d_i} \tag{11-13}$$

$$p_i = \left(\frac{R - d_i}{d_i}\right)^2 \tag{11-14}$$

$$p_i = e^{-\frac{d_i^2}{K^2}} \quad （K\ 为待选常数） \tag{11-15}$$

这 3 种权函数的效果图见图 11-4。

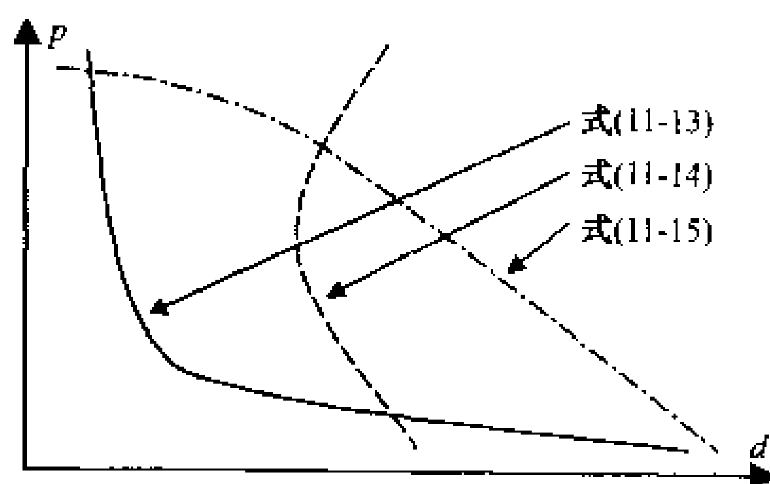


图 11-4 采样点权重与距离的关系示意图

由于 $\overline{x_p} = 0$，$\overline{y_p} = 0$，所以所求得的系数 b_0 即为 P 点高程值。

Hannover 大学的 TASH 程序使用的就是二次曲面移动内插法。值得注意的是：由于解的稳定性决定法方程的状态，面法方程的状态又与点位分布有关，所以当地形起伏较大时，半径 R 不能太大。尤其是当数据较稀疏或分布不均匀时，二次多项式移动内插法容易产生很大的误差。此时，可以考虑采用移动平均法。

3. 移动平均法

移动平均法是一种简化的逐点内插法。其实质是：首先以内插点为中心，确定一个取样窗口，然后计算落在窗口内的采样点的特征观测值的平均值，作为内插点的特征值

估值，如图 11-5 所示。

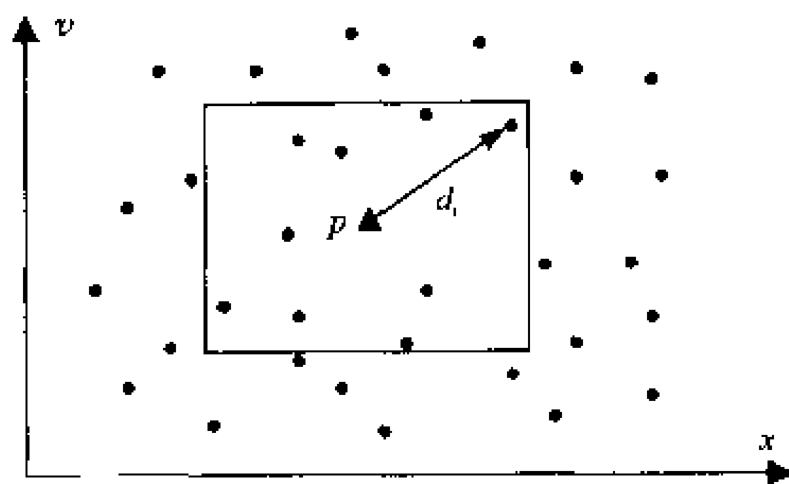


图 11-5 移动平均法取样

对取样窗口的要求是：

(1) 窗口大小要覆盖局域的极大或极小值，以使计算效率与计算精度之间达到合理的均衡；

(2) 窗口内有 4～12 个采样点，即所取采样点数应考虑采样点的分布情况：若规则分布，采样点数可以少些；若非规则分布，采样点应多些。

如 Vienna 工业大学的 SORA 程序则采用了多个邻近点的加权平均水平面进行内插：

$$z_p = \frac{\sum_{i=1}^{n} p_i z_i}{\sum_{i=1}^{n} p_i} \tag{11-16}$$

式中：z_p 为内插点的高程值；z_i 为采样点的高程值；n 为邻近的采样点数目；p_i 为采样点的距离权，可以采用式（11-13）、式（11-14）或式（11-15）求权。

4. 样条函数法

所谓样条函数，即三次多项式。样条函数法的实质为采用三次多项式对采样曲线进行分段修匀。每次的分段拟合仅利用少数采样点的观测值，并要求保持各分段的连接处连续，即光滑可导。其拟合过程相当于用灵曲板来绘制分段连续的曲线。样条函数的一般形式为：

$$f(x,y) = \sum_{r+s=0}^{r+s=3} b_{rs} x^r y^s \tag{11-17}$$

样条函数拟合必须满足观测值与拟合值之差的平方和最小：

$$\sum_{i=1}^{i=n} W_i^2 [z(x_i, y_i) - f(x_i, y_i)]^2 = \min \tag{11-18}$$

式中：W_i 为拟合权，与（x_i，y_i）点处拟合误差的方差成反比；z（x_i，y_i）为（x_i，y_i）点处的采样值；n 为总采样数。

需要 9 个以上的采样点，通过多重回归技术确定样条函数的系数，进而建立基于样条函数的拟合曲线方程。再输入内插点的坐标值（x_u，y_u），就可以解算出内插点的特征值。

5. 多面函数法

多面函数法内插也称多面函数最小二乘法推估，是美国 Hardy 教授于 1977 年提出的。它是从几何观点出发，使采样点形成一个平差的数学曲面问题。其理论依据是分段光滑曲线在三维空间域的扩展："任何一个圆滑的数学表面总是可以用一系列规则的数学表面之和，以任意的精度进行逼近"。也就是说：一个数学表面上某点（x，y）处的高程 z 的表达式为：

$$z = f(x, y) = \sum_{j=1}^{n} a_j q(x, y, x_j, y_j) \tag{11-19}$$

其中 $q(x, y, x_j, y_j)$称为核函数(Kernel)。

若有 $m \geqslant n$ 个采样点,则可以任意选择其中的 n 个为核函数的中心点 $P_j(x_j, y_j)$,并令：

$$q_{ij} = q(x_i, y_i, x_j, y_j) \tag{11-20}$$

则各数据点满足：

$$z_i = \sum_{j=1}^{n} a_j q_{ij} (i = 1, 2, \cdots, m) \tag{11-21}$$

由此可以列出误差方程：

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} - \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = Qa - z \tag{11-22}$$

组成法方程求解得：

$$a = (Q^T Q)^{-1} Q^T z \tag{11-23}$$

于是,任意一点 $P_k(x_k, y_k)$的高程 $z_k(k > n)$为：

$$z_k = Q_k^T \cdot a = Q_k^T (Q^T Q)^{-1} Q^T z \tag{11-24}$$

其中：

$$Q_k^T = [q_{k1} q_{k2}, \cdots, q_{kn}]$$

$$q_{kj} = q(x_k, y_k, x_j, y_j)$$

若将全部数据点取为核函数的中心,即 $n = m$,则：

$$a = Q^{-1} z \tag{11-25}$$

$$z_k = Q_k^T Q^{-1} z \tag{11-26}$$

展开得：

$$z_k = [q_{k1} q_{k2}, \cdots, q_{kn}] \times \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{bmatrix}^{-1} \times \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} \tag{11-27}$$

核函数可以任意选定。通常,可以假定各核函数是对称的圆锥面,如图 11-6 所示。其数学函数式为：

$$q(x,y,x_j,y_j) = \sqrt{(x-x_j)^2+(y-y_j)^2} \tag{11-28}$$

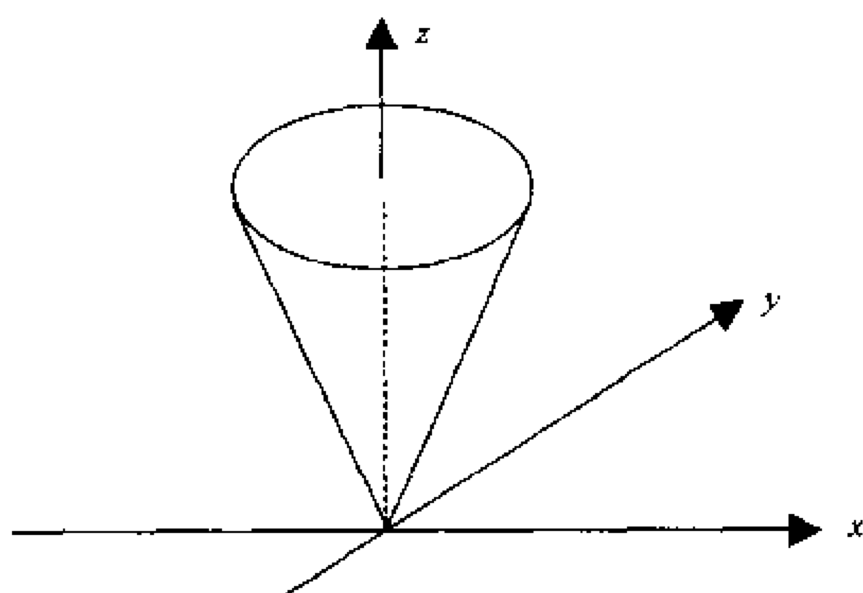


图 11-6 圆锥面核函数

也可以在上式中加入一个常数项 δ，成为一个双曲面，它在采样点处保证坡度的连续性。并令 $d_j=\sqrt{(x-x_j)^2+(y-y_j)^2}$，则上式成为

$$q(x,y,x_j,y_j) = q(d_j) = \sqrt{d_j^2+\delta} \tag{11-29}$$

其他可选的核函数还有：

$$q(d_j) = e^{kd_j^2} \tag{11-30}$$

$$q(d_j) = a^{d_j^2} = 0.995^{d_j^2} \tag{11-31}$$

$$q(d_j) = \frac{1}{1+\left(\frac{d_j}{k}\right)^2} \tag{11-32}$$

$$q(d_j) = d_j^3 + 1 \tag{11-33}$$

$$q(d_j) = \sum_{k=0}^{3} b_k d_j^k \tag{11-34}$$

$$q(d_j) = \sum_{k=0}^{6} b_k d_j^k \tag{11-35}$$

$$q(d_j) = 1 - \frac{d_j^2}{[\max(d_j)]^2} \tag{11-36}$$

$$q(d_j) = e^{-2.5\frac{d_j^2}{[ave(d_j)]^2}} \tag{11-37}$$

其中，最后两个公式是 Arthur 分别于 1965 年和 1973 年所采用的。应该说，Arthur 提出的运算方法是多面函数法最早的一个，其计算公式为：

$$z = k_1q(d_1)+k_2q(d_2)+\cdots+k_nq(d_n) \tag{11-38}$$

6. 线性内插法

当采样点的特征观测值在一个平面上变化时，可以进行线性内插。典型应用是不规则三角高程的内插。认为在一个三角形内，所有点的高程均位于此三角平面上。因此，可

以在该三角形内按直线比例进行线性内插，即可以得到内插点的高程值。其内插函数即为三角形的平面方程：

$$f(x,y) = Ax + By + C \tag{11-39}$$

7. 双线性内插法

分块插区中，当采样点的特征值在 x、y 方向分别按线性规律变化时，需要取按双线性插值法估算内插点的特征值。双线性插值函数为：

$$f(x,y) = Ax + By + Cxy + D \tag{11-40}$$

需要 4 个已知采样点来确定上式中的 4 个系数。4 个已知采样点的选择，如图 11-7 所示，有以下要求：

(1) 环绕内插点，即尽量以内插点为中心均匀分布；

(2) 离内插点距离最近。

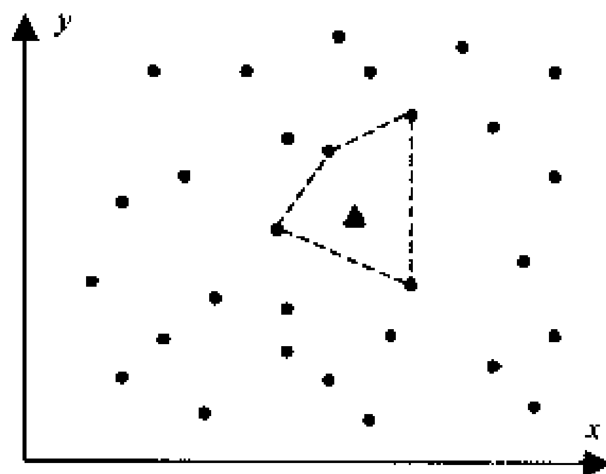


图 11-7　双线性内插的采样点选择

§11.2　区域内插算法

有些采样数据不是均匀变化的，某些数据经专业处理后表达为在各个不同分区中同质或线性变化，这就导致区域之间值的变化在相邻边界处不连续。这种情况有时既不符合现实情况，也不满足连续光滑的假设条件。因而，需要对这些数据进行区域内插，使得数据在边界处连续。区域内插包括点在区域内的内插和面的区域内插两种类型。

11.2.1　点的区域内插算法

已知某种特征数据的分区情况，要求解某待定点的特征值。首先要确定该点位于哪一分区中，然后根据该分区内特征值或其变化规律，内插确定待定点的特征值。

所用算法包括两部分，一是点在多边形内的判断；二是点在区域内的线性插值。其中点在多边形内的判断算法参见第 21 章；点在区域内的线性内插可参考上节线性内插法。

11.2.2 面的区域内插算法

面区域内插的实质是根据某一地区的一组已知分区（称为源区，source region）的已知数据，推求同一分区（目标区，target region）的另一组分区的未知数据。这是GIS的一个非常有用的内插方法，在区域统计、数据挖掘方面有广泛的用途。通常有两种方法，即叠置法和比重法。

1. 叠置法

如图11-8所示，设源区为3个县的人口统计数据，其统计结果见表11-2；图11-9所示是目标区为按流域分的3个自然地理区块。现要研究人口密度与地理区块之间的关系。过程如下：

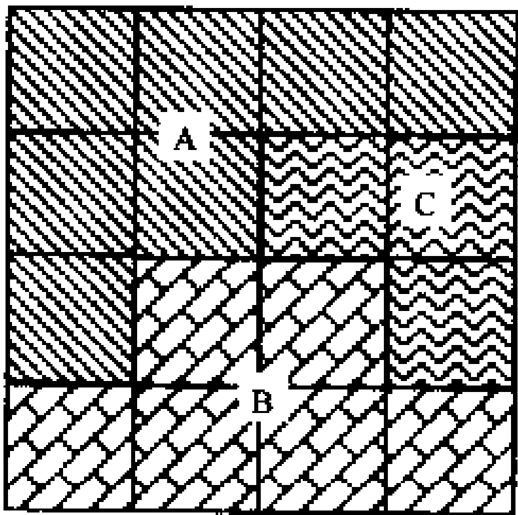


图11-8 源区分区情况

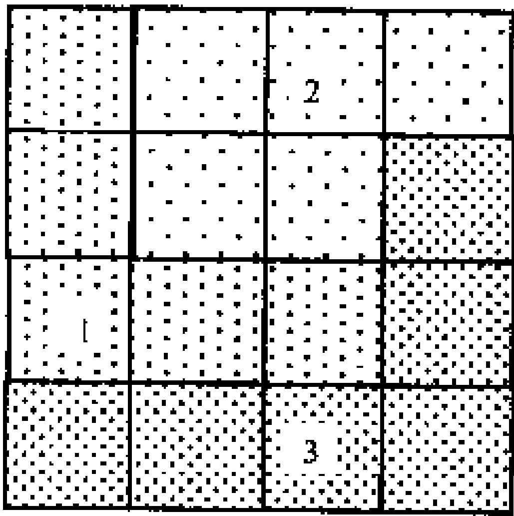


图11-9 目标区分区情况

表11-2 源区人口统计

县名	A	B	C
土地面积（km^2）	7000	6000	3000
人口数（千人）	350	300	100

第一步：将目标区叠加在源区上，计算两者的交集的面积 a_{st}，结果见表11-3；

表11-3 源区与目标区的交集

交集 源区 / 目标区	A	B	C
1	3000	2000	0
2	4000	0	1000
3	0	4000	2000

第二步：按下式计算目标区中各分区的内插人口值：

$$V_t = A_{ts}(A_s)^{-1}U_s \tag{11-41}$$

式中：

$V_t = \begin{bmatrix} v_{t1} \\ v_{t2} \\ \vdots \\ v_{tm} \end{bmatrix}$；为目标区的分区人口矩阵；

$A_{ts} = \begin{bmatrix} a_{t_1s_1} & a_{t_1s_2} & \cdots & a_{t_1s_n} \\ a_{t_2s_1} & a_{t_2s_2} & \cdots & a_{t_2s_n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{t_ms_1} & a_{t_ms_2} & \cdots & a_{t_ms_n} \end{bmatrix}$，为目标区与源区交集的面积矩阵；

$A_s = \begin{bmatrix} A_{s1} \\ A_{s2} \\ \vdots \\ A_{sn} \end{bmatrix}$，为源区的分区面积矩阵；

$U_s = \begin{bmatrix} u_{s1} \\ u_{s2} \\ \vdots \\ u_{sn} \end{bmatrix}$，为源区的分区人口矩阵。

插值计算结果见表 11-4。

表 11-4　目标区的人口插值结果

自然地理分区	1	2	3
土地面积（km^2）	5000	5000	6000
人口数（千人）	250	233.333	266.667

上述公式的应用条件是：已知源区中各分区的某项绝对统计值，如人口总数、资源总数、GDP、总收入等。如果已知的不是绝对数值，而是源区中各分区的某项密度数据，如人口密度、单位面积资源量等，则计算更为简单：

$$V_t = A_{ts}P_s \tag{11-42}$$

式中：$P_s = \begin{bmatrix} p_{s1} \\ p_{s2} \\ \vdots \\ p_{sn} \end{bmatrix}$，为源区的分区密度矩阵。

得到目标分区的绝对统计值后，再除以目标分区的面积，还可以得到目标分区的密度统计数据如下：

$$P_t = V_t(A_t)^{-1} \tag{11-43}$$

式中：$A_t = \begin{bmatrix} a_{t1} \\ a_{t2} \\ \vdots \\ a_{tm} \end{bmatrix}$，为目标区的分区面积矩阵。

有一类特殊情况需要注意。当源区的统计数据为比值时，如人均 GDP、人均资源占有量、单位建筑面积绿化率等，其内插问题就变成一个二次内插计算。

二次内插的步骤如下：

(1) 一次内插：首先，以比值的分母量（如人口总数、建筑总面积）为中间参数，建立与源区内各分区对应的中间参数数据表（必须是已知的），并按公式（11-41）内插出目标区内各分区的中间参数数据表；

(2) 二次内插：然后，用源区内各分区的中间参数数据乘以相应的比值统计数据，得到源区内各分区的绝对统计数据；就可以按公式（11-41）计算目标区内各分区的绝对统计数据；最后，用目标区内各分区的绝对统计数据除以目标区内各分区的中间参数数据，就可以得到目标区内各分区的比值数据。

2. 比重法

所谓比重法，是根据平滑密度函数的原理，将源区内的统计数据从各分区内的均匀分布转变为分区内的非均匀分布，而非均匀分布更代表社会经济和自然现象的客观规律。比如：

(1) 人们通常用一个县气象站的气象数据代表该县范围内的气象情况；邻县气象站的气象数据代表邻县范围内的气象情况。事实上，两县的交界地区的气象情况比较复杂，应具有两县气象站的气象数据的综合特征；

(2) 人们在处理地质调查数据时，通常用某一位置的调查数据代表某区的平均指标，而形成某种区内均匀、区间不连续的分区报告。事实上，在区域的交界处，地质调查参数往往是连续渐变的（断层影响除外）。

针对以上问题，可以根据平滑密度函数的原理，来对各类分区突变数据进行渐变处理，过程如下：

(1) 采用某种栅格尺寸将原图栅格化，栅格尺寸的大小应保证满足内插精度要求；

(2) 根据原图各分区的面积和栅格尺寸，将分区统计数据按比例赋予其中的每一个栅格，并计算全图的统计数据总和 U；

(3) 按以下公式计算每一栅格的上、下、左、右 4 个邻近栅格的数据的平均值，并作为该栅格的新值刷新它；

$$z_{i,j} = (z_{i,j-1} + z_{i,j+1} + z_{i-1,j} + z_{i+1,j})/4 \tag{11-44}$$

(4) 计算所有栅格刷新后全图的统计数据总和 U'；

(5) 计算系数 $p = U/U'$，将刷新后全图的每一栅格均乘以该系数，得到又一次刷新的新图；

(6) 重复第 3～5 步，依次进行下去，直到 p 趋近于于 1 为止。

图 11-10 所示为比重法区域内插的一个简单实例。其中 (a) 为原始数据；(b) 为按式（11-44）进行的第一次刷新；(c) 为乘以系数 p 后的刷新；(d) 为最后结果。

5.0	5.0	5.0	5.0
5.0	5.0	3.3	3.3
5.0	5.0	5.0	3.3
5.0	5.0	5.0	5.0

(a)

5.0	5.0	4.4	4.2
5.0	4.6	4.6	3.9
5.0	5.0	4.2	4.4
5.0	5.0	5.0	4.2

(b)

5.3	5.3	4.6	4.4
5.3	4.8	3.6	3.0
5.3	5.3	4.5	3.4
5.3	5.3	5.3	4.5

(c)

5.1	5.1	4.6	4.0
5.3	5.1	3.5	3.2
5.5	5.2	4.5	3.3
5.5	5.5	4.9	4.5

(d)

图 11-10 比重法区域内插实例

§11.3 采样点曲线拟合

GIS 在许多情况下需要将已知的离散采样点值连接成光滑曲线，如通过地面水准测量点、GPS 测高点等勾画等高线；通过手工数字化采集等高线数据后，重新内插生成等高线；以及通过离散的实测或统计采样点，生成等值区域等。这一过程的实质就是采样点的曲线光滑处理，也称采样点曲线拟合。采样点曲线拟合的方法有张力样条函数、分段三次多项式、分段圆弧和切线方向内插等。

11.3.1 张力样条函数法

如图 11-11 所示，已知平面上的非等距采样点 p_1（x_1，y_1），p_2（x_2，y_2），…，p_n（x_n，y_n），从第一点开始其累加弦长为：

$$s_{i+1} = s_i + \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{11-45}$$

满足 $s_1 < s_2 < \cdots < s_n$。

选择二阶导数连续的单值函数 $x = x(s)$ 和 $y = y(s)$，使之满足：

$$\left.\begin{aligned} x_i &= x(s_i) \\ y_i &= y(s_i) \end{aligned}\right\} \quad (i = 1, 2, \cdots, n)$$

再给定一个常数 $\sigma \neq 0$，使得 $x''(s) - \sigma^2 x(s)$ 和 $y''(s) - \sigma^2 y(s)$ 都在区间 $[s_i, s_{i+1}]$ 上呈线性变化，即：

$$x''(s) - \sigma^2 x(s) = [x''(s_i) - \sigma^2 x_i] \times \frac{s_{i+1} - s}{h_i} + [x''(s_{i+1}) - \sigma^2 x_{i+1}] \times \frac{s - s_i}{h_i} \tag{11-46}$$

$$y''(s) - \sigma^2 y(s) = [y''(s_i) - \sigma^2 y_i] \times \frac{s_{i+1} - s}{h_i} + [y''(s_{i+1}) - \sigma^2 y_{i+1}] \times \frac{s - s_i}{h_i} \tag{11-47}$$

其中 $h_i = s_{i+1} - s_i$，为相邻采样点之弦长。

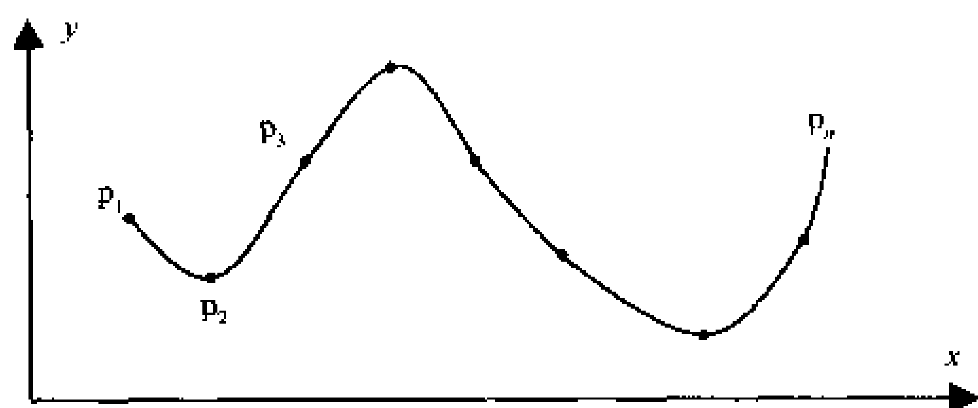


图 11-11　张力样条函数拟合示意图

以上两式的解函数即为所要选择的二阶导数连续的单值函数，称为张力样条函数，记为：

$$\left.\begin{aligned}
x(s) &= \frac{1}{\sigma^2 sh(\sigma h_i)}\{x''(s_i)sh[\sigma(s_{i+1} - s)] + x''(s_{i+1})sh[\sigma(s - s_i)]\} + \\
&\quad \left[x_i - \frac{x''(s_i)}{\sigma^2}\right]\times\frac{s_{i+1} - s}{h_i} + \left[x_{i+1} - \frac{x''(s_{i+1})}{\sigma^2}\right]\times\frac{s - s_i}{h_i} \\
y(s) &= \frac{1}{\sigma^2 sh(\sigma h_i)}\{y''(s_i)sh[\sigma(s_{i+1} - s)] + y''(s_{i+1})sh[\sigma(s - s_i)]\} + \\
&\quad \left[y_i - \frac{y''(s_i)}{\sigma^2}\right]\times\frac{s_{i+1} - s}{h_i} + \left[y_{i+1} - \frac{y''(s_{i+1})}{\sigma^2}\right]\times\frac{s - s_i}{h_i} \\
s_i &\leqslant s \leqslant s_{i+1} \quad (i = 1,2,\cdots,n-1)
\end{aligned}\right\} \tag{11-48}$$

其中 σ 称为张力系数。通过改变 σ 的大小，可以调节拟合曲线的光滑程度：σ 趋近于 0，则曲线为三次样条；当 σ 趋近于无穷大，则曲线退化为折线。张力样条法插值的缺点是计算量大。

11.3.2　分段三次多项式法

分段三次多项式法又称五点光滑法，其原理是在每两个采样点之间建立一条三次多项式曲线方程，且整条曲线具有连续的一阶导数；各采样点的一阶导数以该点及其两侧相邻各两点（共 5 点）来确定，从而保证曲线是光滑的。如图 11-12 所示，已知平面上有若干采样点 p_{i-2} (x_{i-2}, y_{i-2})，p_{i-1} (x_{i-1}, y_{i-1})，p_i (x_i, y_i)，p_{i+1} (x_{i+1}, y_{i+1})，p_{i+2} (x_{i+2}, y_{i+2})，在任意两个相邻采样点之间，用一条三次曲线 $f(x)$ 进

行拟合，并且要求拟合曲线严格通过该两点，同时要求它在这两个已知点处的一阶导数等于给定值。

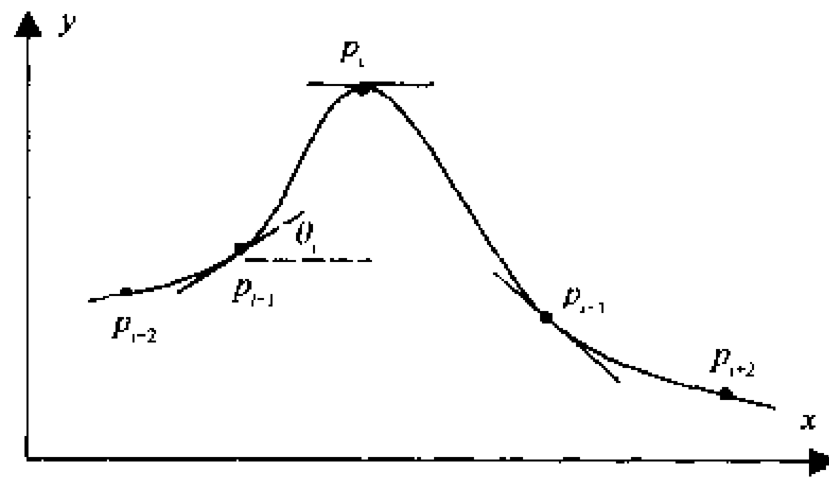


图 11-12　分段多项式拟合示意图

设相邻两点 p_i、p_{i+1}的连线线段 p_ip_{i+1}的斜率为 k_i，则拟合后两点处的斜率应为：

$$
\begin{aligned}
\tan\theta_i &= \frac{|k_{i+1}-k_i|\cdot k_{i-1}+|k_{i-1}-k_{i-2}|\cdot k_i}{|k_{i+1}-k_i|+|k_{i-1}-k_{i-2}|} \\
\tan\theta_{i+1} &= \frac{|k_{i+2}-k_{i+1}|\cdot k_i+|k_i-k_{i-1}|\cdot k_{i+1}}{|k_{i+2}-k_{i+1}|+|k_i-k_{i-1}|}
\end{aligned}
\tag{11-49}
$$

设过相邻两点的拟合曲线方程为：

$$y = f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \tag{11-50}$$

该拟合曲线应满足以下 4 个条件：

$$y_i = f(x_i);\qquad y_{i+1} = f(x_{i+1})$$

$$y'_i = f'(x_i) = \tan\theta_i;\qquad y'_{i+1} = f'(x_{i+1}) = \tan\theta_{i+1}$$

根据以上 4 个等式，即能解算出式（11-50）中的四个待定系数：

$$
\left.
\begin{aligned}
a_0 &= y_i \\
a_1 &= r\sin\theta_i \\
a_2 &= 3(y_{i+1}-y_i) - r(\sin\theta_{i+1} + 2\sin\theta_i) \\
a_3 &= -2(y_{i+1}-y_i) + r(\sin\theta_{i+1} + \sin\theta_i) \\
r &= \sqrt{(x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2}
\end{aligned}
\right\}
\tag{11-51}
$$

由于分段三次多项式拟合是每相邻两个点（x_{i-1}，y_{i-1}）和（x_i，y_i）；（x_i，y_i）和（x_{i+1}，y_{i+1}），（x_{i+1}，y_{i+1}）和（x_{i+2}，y_{i+2}）依次进行的，因此可以满足整个曲线光滑连续的条件。

需要指出的是，在拟合曲线的前后 4 个端点处（即最前面的两点和最后面的两点）为了利用公式（11-49）计算端点处的曲线斜率并保证曲线在首尾端的连续性，必须按一定的间距规则分别补两个点，使得所补的两点与曲线的首（或尾）3 点位于同一抛物线上。设拟合曲线最末尾的连续 3 点分别为 p_i（x_i，y_i），p_{i+1}（x_{i+1}，y_{i+1}），p_{i+2}（x_{i+2}，y_{i+2}），所补的两点依次为 p_{i+3}（x_{i+3}，y_{i+3}），p_{i+4}（x_{i+4}，y_{i+4}），且满足：

$$x_{i+4} - x_{i+2} = x_{i+3} - x_{i+1} = x_{i+2} - x_i$$

$$k_{i+3} - k_{i+2} = k_{i+2} - k_{i+1} = k_{i+1} - k_i$$

11.3.3 分段圆弧法

如图 11-13 所示，已知平面上有若干采样点 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n)，在任意两个相邻采样点 $p_i(x_i, y_i)$和 $p_{i+1}(x_{i+1}, y_{i+1})$之间，用两条相切圆弧进行拟合，并且要求拟合曲线严格通过该两点，同时要求它在这两个已知点处的一阶导数也等于给定值。

显然，根据上述规则，一般可以得出多组满足条件的圆弧。若再加以一定的条件限制，则可以得到惟一解。

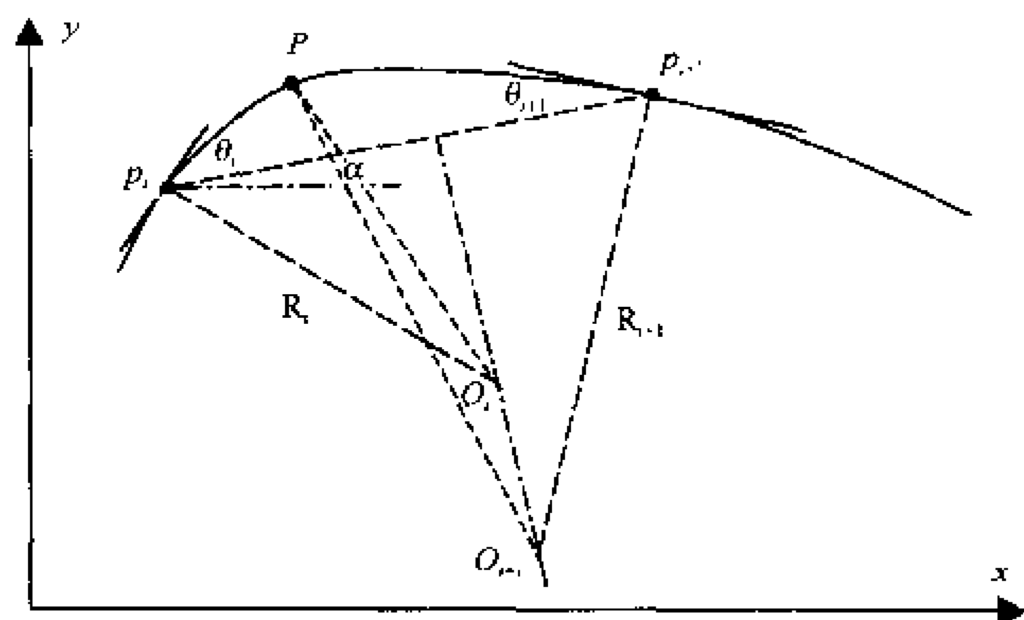


图 11-13　限制型分段圆弧拟合示意图

例如，假设两个采样点的连线 p_ip_{i+1}与两个拟合圆的圆心的连线（O_iO_{i+1}）互相垂直，则有：

$$\tan\alpha = (y_{i+1} - y_i)/(x_{i+1} - x_i)$$

$$d = p_ip_{i+1} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

$$\Delta\alpha_i = |\alpha - \theta_i| = \angle p_iO_iP$$

$$\Delta\alpha_{i+1} = |\alpha - \theta_{i+1}| = \angle p_{i+1}O_{i+1}P$$

式中：α 为 p_ip_{i+1}与 x 轴的夹角；θ_i 为过 p_i 点的切线与 p_ip_{i+1}的夹角；θ_{i+1}为过 p_{i+1}点的切线与 $p_{i+1}p_i$ 的夹角。

根据以下条件：

$$d = p_ip_{i+1} = r_i\sin\Delta\alpha_i + r_{i+1}\sin\Delta\alpha_{i+1}$$

$$r_i - r_i\cos\Delta\alpha_i = r_{i+1} - r_{i+1}\cos\Delta\alpha_{i+1}$$

即可解得：

$$r_i = d\Big/\left(\sin\Delta\alpha_i + \frac{1 - \cos\Delta\alpha_i}{1 - \cos\Delta\alpha_{i+1}} \times \sin\Delta\alpha_2\right) \tag{11-52}$$

$$r_{i+1} = r_i \times \frac{1 - \cos\Delta\alpha_i}{1 - \cos\Delta\alpha_{i+1}} \tag{11-53}$$

进而可以解算出 O_i、O_{i+1}的坐标。

11.3.4 切线方向法

也可以根据采样点处的切线方向来拟合采样曲线，进而进行空间数据内插。根据内插时使用采样点的数量，又分为三点法和五点法。

1. 三点法

即假设某一采样点 p_i（x_i，y_i）上的切线垂直于该点的两个相邻点：p_{i-1}（x_{i-1}，y_{i-1}）与 p_{i+1}（x_{i+1}，y_{i+1}）的张角的角平分线，即点 p_i 处的切线方向角为：

$$\theta_i = \frac{\pi}{2} + \frac{1}{2}\left(\arctan\frac{y_{i+1} - y_i}{x_{i+1} - x_i} + \arctan\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right) \tag{11-54}$$

另一种方法是：假设 p_i（x_i，y_i）点的切线方向与该点的两个相邻点：p_{i-1}（x_{i-1}，y_{i-1}）与 p_{i+1}（x_{i+1}，y_{i+1}）的连线方向一致，即：

$$\theta_i = \arctan\frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}} \tag{11-55}$$

2. 五点法（Akima 法）

Akima 法由 5 个相邻点 p_k（x_k，y_k）（$k = i-2$，$i-1$，i，$i+1$，$i+2$）解算曲线在 p_i 点的斜率，它是以 p_i 为端点的两弦斜率的加权平均值，其权 P_r 与 P_l 分别等于 p_i 点前两弦斜率差的绝对值与后两弦斜率差的绝对值：

$$\tan\theta_i = \frac{P_l\tan\frac{y_{i+1} - y_i}{x_{i+1} - x_i} + P_r\tan\frac{y_i - y_{i-1}}{x_i - x_{i-1}}}{P_l + P_r} \tag{11-56}$$

其中：

$$P_l = \left|\tan\frac{y_i - y_{i-1}}{x_i - x_{i-1}} - \tan\frac{y_{i-1} - y_{i-2}}{x_{i-1} - x_{i-2}}\right|$$

$$P_r = \left|\tan\frac{y_{i+2} - y_{i+1}}{x_{i+2} - x_{i+1}} - \tan\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right|$$

实践表明，Akima 法最接近训练有素的绘图员徒手绘制的内插曲线，尤其当有三个点位于一条直线上时，Akima 法能内插出一条直线，而其他非线性内插法则没有这种特性。Akima 法的缺点是，采样点处的二阶导数不能保证连续。

除以上介绍的采样点曲线拟合算法之外，还有线性迭代（抹角法）、二次多项式平均加权法（正轴抛物线平均加权法）和斜轴抛物线平均加权法等，刘岳等（1981）对此有详细介绍，并基于 6 种采样条件对这些曲线的光滑效果进行了比较。

参 考 文 献

龚健雅. 2001. 地理信息系统基础. 北京：科学出版社. 174～189

李庆扬，王能超，易大义. 1982. 数值分析. 武汉：华中理工大学出版社. 19～60

刘纯波，吴立新，陈桂茹，崔志琳. 1998. 一种基于空间边界控制与八叉树模型的空间插值方法. 中国煤炭学会第五届青年科技学术年会. 北京：煤炭工业出版社：350～355

刘　岳，梁启章．专题地图制图自动化．北京：测绘出版社．91～121
沈永欢，梁在中，许履瑚，蔡倩倩．2002．数学手册．北京：科学出版社
於宗铸，鲁林成．1983．测量平差基础．北京：测绘出版社
Kridge D G., 1996. A practical analysis of the effects of spatial structures and data available and accessed on conditional bias in ordinary Kriging. Proc. 5th Int. Geostatistics Congress. Wollongong, Australia
Krige D. G., 1999. Conditional biases and uncertainty of estimation in geostatistics. Proc. APCOM' 99 Symposium, Colorado School of Mines
Krige D. G. and Assibey A. B., 2001. Valuation of revoverable resource by Kriging, directi condition or simulation. Proc. APCOM' 01 Symposium, China Uni. Mining and Tech
Simon W H., 1994. 3D Geoscience Modeling: Computer Techniques for Geological Characterization. Berlin: Springer-Verlag
Simon W. H. 2000. Practical Geostatistics: Modeling and Spatial Analysis. Berlin: Springer-Verlag

第十二章　空间数据转换算法

空间数据转换既是 GIS 数据处理的一项重要任务，也是 GIS 的技术难题之一。有时候，为了方便分析和应用，需要将矢量数据转换为栅格数据，或者将栅格数据转换为矢量数据。因为转换程序通常占用较多的内存，涉及复杂的计算，过去一直难以在微机环境下进行。近年，随着计算机处理能力的不断提高，已经发展了许多高效的转换算法。

对于点状实体而言，每个实体仅由一个坐标对表示，其矢量结构和栅格结构的相互转换基本上只是坐标精度转换的问题。线实体的矢量结构在转换为栅格数据时，除了要把矢量坐标转换为栅格行列坐标外，还要根据转换精度要求，在坐标点之间进行栅格内插，这可以由两点式直线方程得到。反之，线实体的栅格结构转换为矢量结构时，与将栅格面域多边形转换为矢量多边形的方法相似。本章将讨论基于弧段和基于面域多边形的矢量结构与栅格结构的相互转换问题。

§12.1　矢量数据向栅格数据转换

矢量结构向栅格结构转换又称为多边形填充，即在矢量表示的多边形边界内部的所有栅格上赋予相应的多边形编号，从而形成栅格数据阵列。算法主要有内部点扩散法、射线算法、扫描算法、复数积分算法和边界代数算法等。

12.1.1　内部点扩散算法

内部点扩散法是基于面域多边形的。如图 12-1 所示，其扩散过程如下：

(1) 按一定栅格尺寸将矢量图经栅格化后，对矢量图内每个面域多边形分别选择一个内部点（种子点）；

(2) 从种子点开始，向其 8 个相邻栅格扩散，分别判断这 8 个栅格是否在多边形的边界上：若是，则该栅格不作为种子点；若不是，则该栅格作为新的种子点；

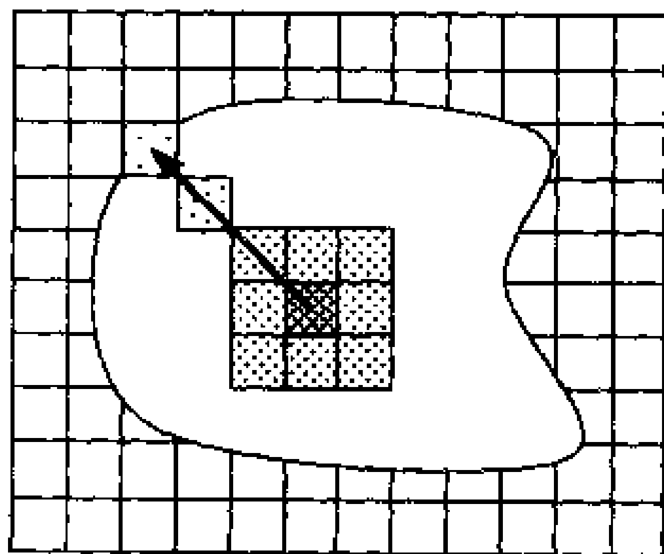


图 12-1　内部点扩散原理

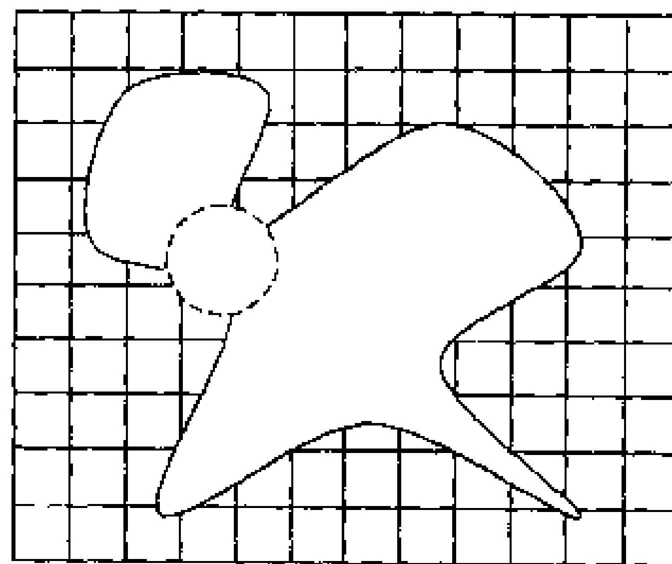


图 12-2　多边形不连通实例

(3) 新种子点与原种子点一起进行新的扩散运算；

(4) 重复以上过程，直到所有新老种子点填满该多边形并遇到边界为止。

该法的缺点是：算法程序设计比较复杂，需要在栅格矩阵中进行搜索，内存消耗较大；当栅格尺寸取得不合理（过大）时，某些复杂图形（如狭长多边形）的两条边界落在同一个或相邻的两个栅格内，会造成多边形不连通，如图 12-2 所示。此时，一个种子点不能完成整个多边形的填充。

12.1.2 射线算法

射线法也是基于多边形的，是矢量条件下点在多边形内部的判断算法（见 14.1 节）的推广。其实质是逐个栅格判断其是否位于某个多边形之内：由待定栅格向任意方向引射线，判断该射线与某多边形所有边界的相交总次数；如果相交偶数次，则待定点在该多边形的外部，如为奇数，则待定点在该多边形内部。如图 12-3 所示。

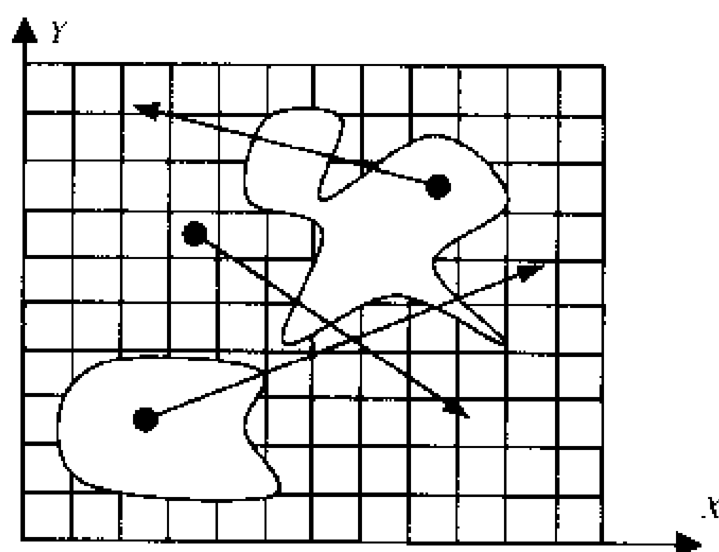


图 12-3　射线法判断原理

射线法要逐个栅格计算其与图内所有多边形的所有交点，因此运算量很大。此外，该算法还存在不完善的地方，即射线与多边形边界可能相切、重合等，这会影响交点数的统计，必须予以判断并排除，从而增加了算法的难度。

12.1.3 扫描算法

根据实施细节不同，扫描算法又分为行扫描算法、扫描线算法和带扫描算法。

1. 行扫描算法

行扫描算法是对射线法的改进，即确定射线方向只沿栅格矩阵的行方向，也称为行扫描。其实施办法是：在行扫描每两次遇到同一多边形的边界时，其间的栅格属于该多边形。可见，扫描算法省去了计算射线与多边形边界交点的大量运算，可以大大提高效率。但需要预留一个较大的数组以存放边界点；而且扫描线与多边形边界可能相切、重合的问题依然存在，还需要加以判断和取舍。

2. 扫描线算法

扫描线算法类似于行扫描算法，也是基于面域多边形的。所不同的是扫描线算法主要依靠数学计算，而不是通过实际的扫描过程来完成栅格所属关系的判断。如图 12-4 所示，对于凸多边形，其原理和算法程序如下：

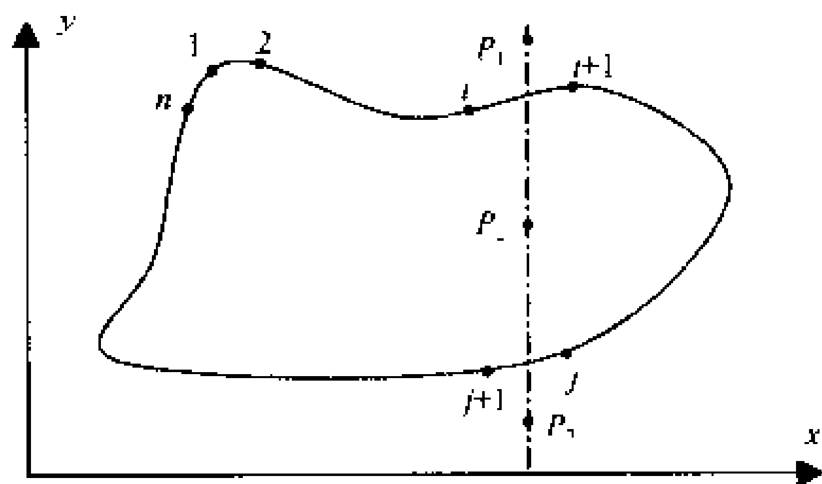


图 12-4　凸多边形扫描线法原理

第一步：提取面域多边形的边界坐标数据文件；

第二步：边界上任选一点开始，分别按顺时针方向和逆时针方向将该面域的所有边界点的 x 坐标排队：x_1，x_2，x_3，…，x_n 或 x_n，x_{n-1}，x_{n-2}，…，x_1；

第三步：根据待定栅格 P（x_p，y_p）的 x 坐标，判断 x_p 分别落在上述两组队列中的区间位置，并找出 P 点所在各区间位置的左右 x 坐标，如顺时针队列中的两个区间依次为 $x_i \sim x_{i+1}$ 和 $x_j \sim x_{j+1}$（$j > i$），逆时针队列中的两个区间依次为 $x_{j+1} \sim x_j$ 和 $x_{i+1} \sim x_i$；

第四步：按线性内插算法，分别计算 $x_i \sim x_{i+1}$（或 $x_{i+1} \sim x_i$）区间和 $x_j \sim x_{j+1}$（或 $x_{j+1} \sim x_j$）区间对应 x_p 的 y_m 值（$m = 1$，2），即 y_1 和 y_2：

$$y_1 = y_i + \frac{x_p - x_i}{x_{i+1} - x_i} \times (y_{i+1} - y_i) \tag{12-1}$$

$$y_2 = y_j + \frac{x_p - x_j}{x_{j-1} - x_j} \times (y_{j-1} - y_j) \tag{12-2}$$

第五步：计算顺、逆时针方向的两组中间参数 A_k 和 B_k（$k = 1$，2，分别对应顺、逆时针方向所得区间的顺序，对于顺时针有 $k = m$，对于逆时针有 $k = 3 - m$）：

$$A_k = \begin{cases} 1 & y_k \geqslant y_p \\ -1 & y_k < y_p \end{cases} \tag{12-3}$$

$$B_k = \begin{cases} 1 & y_k < y_p \\ -1 & y_k \geqslant y_p \end{cases} \tag{12-4}$$

第六步：按下式判断，若 $|C| = A_k + B_k = 2$，则 P 栅格位于本面域多边形内；否则，P 栅格位于本面域多边形之外。

作者将凸多边形的扫描线法作了扩展，可以对任意复杂的非凸多边形进行由矢量到栅格的转换。如图 12-5 所示，扩展后的扫描线法原理和算法程序如下：

第一步：提取面域多边形的边界坐标数据文件；

第二步：边界上任选一点开始，分别按顺时针方向和逆时针方向将该面域的所有边界点的 x 坐标排队：x_1，x_2，x_3，…，x_n 或 x_n，x_{n-1}，x_{n-2}，…，x_1；

第三步：根据待定栅格 P（x_p，y_p）的 x 坐标，判断 x_p 分别落在上述两组队列中的区间位置，并找出 P 点所在各区间位置的左右 x 坐标（$x_i \sim x_{i+1}$），如顺时针队列中

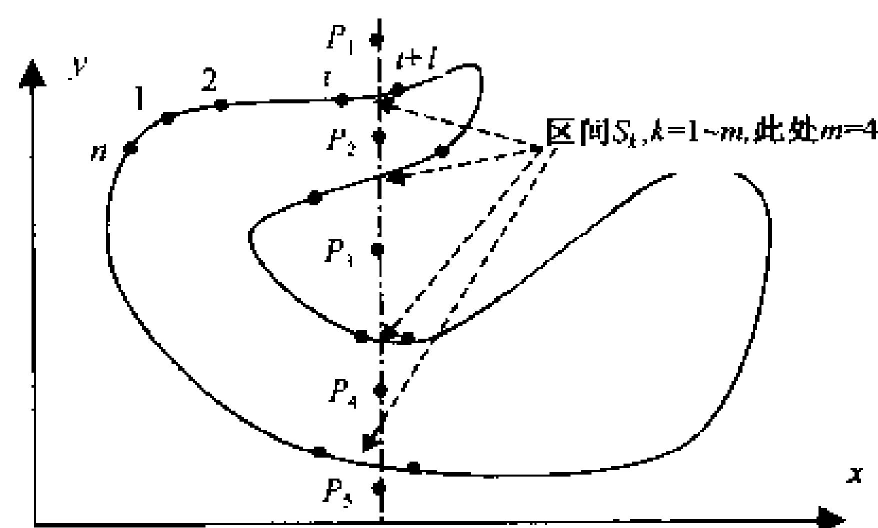


图 12-5　任意非凸多边形扫描线法原理

的第一区间为 S_1，其他依次为 S_2，…，S_m；逆时针队列中的区间依次为 S_m，S_{m-1}，…，S_1；

第四步：按线性内插算法，分别计算各区间内对应 x_p 的 y_k 值（$k=1, 2, \cdots, m$）：

$$y_k = y_i + \frac{x_p - x_i}{x_{i+1} - x_i} \times (y_{i+1} - y_i) \tag{12-5}$$

第五步：计算顺、逆时针方向的两组中间参数 A_k 和 B_k（对于顺时针有 $k=k$，对于逆时针有 $k=m-k+1$，其中 $k=1, 2, \cdots, m$）：

$$A_k = \begin{cases} 1 & y_k \geqslant y_p \\ -1 & y_k < y_p \end{cases} \tag{12-6}$$

$$B_k = \begin{cases} 1 & y_k < y_p \\ -1 & y_k \geqslant y_p \end{cases} \tag{12-7}$$

第六步：若 $C_k = |(A_k + B_k)| = 2$，记数 $N_k = 1$；否则 $N_k = 0$；并累计 $N = \sum_{k=1}^{k=m} N_k$；

第七步：判断：若 $N/2$ 为奇数，则 P 栅格位于本面域多边形内；否则，P 栅格位于本面域多边形之外。

以图 12-5 中的 $P_1 \sim P_5$ 为例，其判断结果如下表所示。

表 12-1　图 12-5 的点在多边形内判断结果

待定点	A_k				B_k				C_k				$N/2$	在否
	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4		
P_1	−1	−1	−1	−1	1	1	1	1	0	0	0	0	0	否
P_2	1	−1	−1	−1	1	1	1	−1	2	0	0	2	1	是
P_3	1	1	−1	−1	1	1	−1	−1	2	2	2	2	2	否
P_4	1	1	1	−1	1	−1	−1	−1	2	0	0	2	1	是
P_5	1	1	1	1	−1	−1	−1	1	0	0	0	0	0	否

以图 12-6 中所示复杂非凸多边形及 $P_1 \sim P_{10}$ 为例，其判断结果如表 12-2 所示：

表 12-2　图 12-6 点在多边形内判断结果

待定点		P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}
A_k	S_1	-1	1	1	1	1	1	1	1	1	1	1
	S_2	-1	-1	1	1	1	1	1	1	1	1	1
	S_3	1	-1	-1	1	1	1	1	1	1	1	1
	S_4	-1	-1	-1	-1	-1	-1	1	1	1	1	1
	S_5	-1	-1	-1	1	-1	-1	-1	-1	-1	1	1
	S_6	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1
	S_7	-1	-1	-1	-1	-1	1	-1	1	1	1	1
	S_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
	S_9	-1	-1	-1	1	-1	1	1	1	1	1	1
	S_{10}	-1	-1	-1	-1	1	1	1	1	1	1	1
B_k	S_1	1	1	1	1	-1	-1	-1	1	-1	-1	-1
	S_2	1	1	1	1	1	-1	-1	1	-1	-1	-1
	S_3	1	1	1	1	1	1	1	1	1	1	-1
	S_4	1	1	1	1	1	1	1	-1	-1	-1	-1
	S_5	1	1	1	1	1	1	1	1	-1	-1	-1
	S_6	1	1	1	1	1	1	1	1	1	1	-1
	S_7	1	1	1	1	1	1	-1	-1	-1	-1	-1
	S_8	1	1	1	-1	1	-1	-1	-1	1	-1	-1
	S_9	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1
	S_{10}	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
C_k	S_1	0	2	2	2	0	0	0	0	0	0	0
	S_2	0	0	2	2	2	0	0	0	0	0	0
	S_3	0	0	0	2	2	2	2	2	2	2	0
	S_4	0	0	0	0	0	0	2	0	0	0	0
	S_5	0	0	0	0	0	0	0	0	2	0	0
	S_6	0	0	0	0	0	0	0	0	2	0	0
	S_7	0	0	0	0	0	0	2	0	0	0	0
	S_8	0	0	0	2	2	2	2	2	2	2	0
	S_9	0	0	2	2	2	0	0	0	0	0	0
	S_{10}	0	2	2	2	0	0	0	0	0	0	0
$N/2$		0	1	2	3	2	1	2	1	2	1	0
在否		否	是	否	是	否	是	否	是	否	是	否

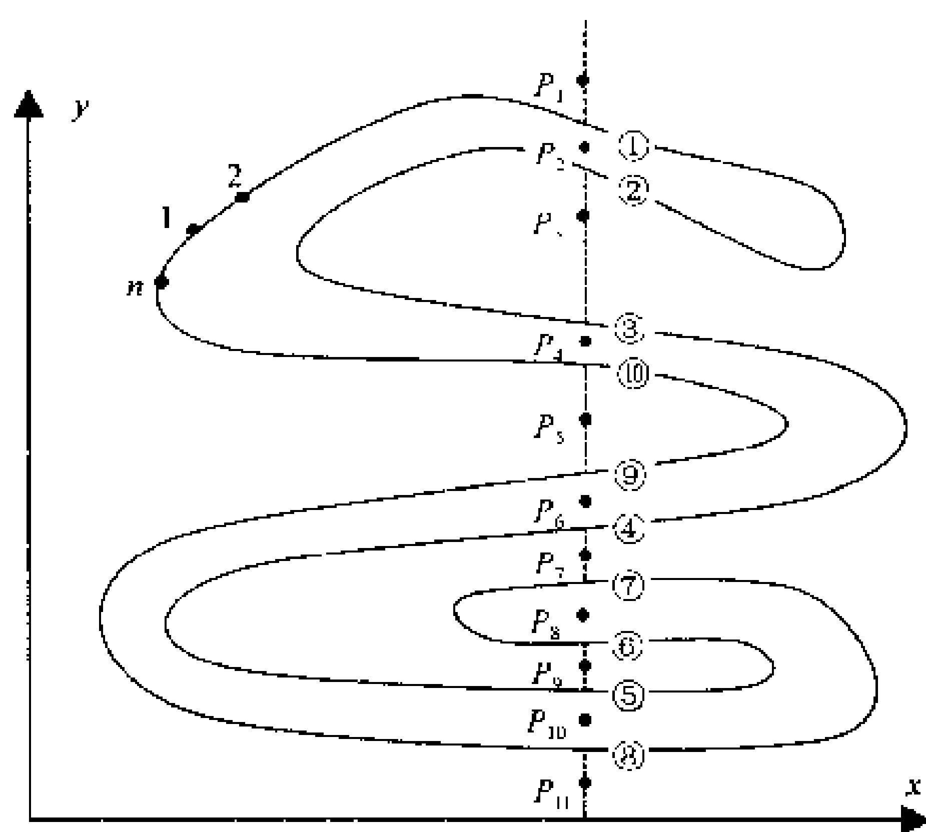


图 12-6 复杂非凸多边形扫描线法实例

3. 带扫描算法

带扫描算法是基于弧段，在计算机内存容量有限而需进行大数据量转换时采用的。其基本思想是：首先按一定带宽将矢量地图分为若干带（如图 12-7 所示），然后对全图进行一次总体扫描，确定该图范围内的几个关键参数：图幅角点坐标、分带数 N_w、栅格行数 N_r、栅格列数 N_c、弧段总数 N_a 等，进而逐带、逐弧进行扫描和栅格归属判断。扫描流程如图 12-8 所示。

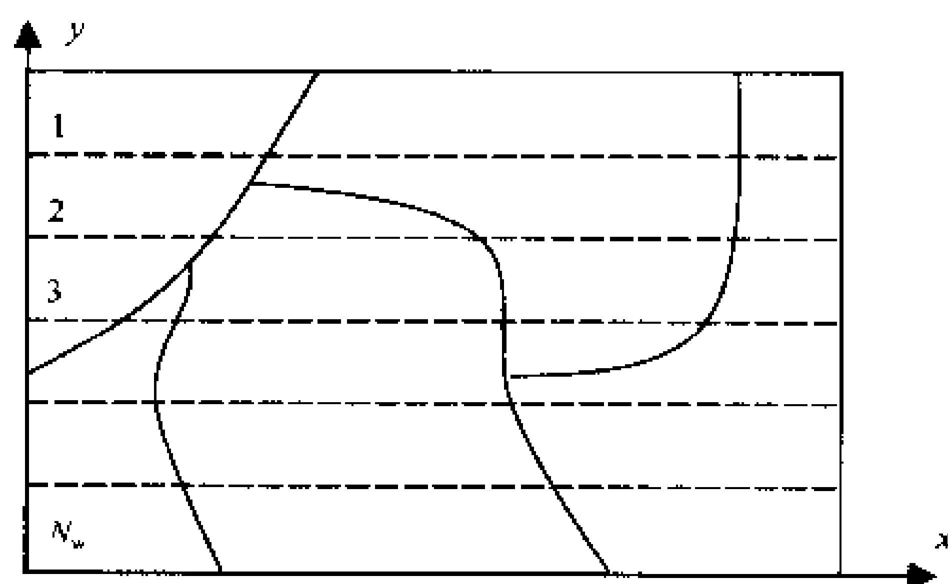


图 12-7 带扫描算法实例

带扫描算法的关键是要记录扫描线与弧段之交点的左右属性编码，然后根据在同一行中按 x 坐标由小到大的顺序和所记录的左右属性码依次进行配对。这一过程，其实就是确定每一行的栅格游程，是栅格数据压缩算法中的游程编码技术的实用。

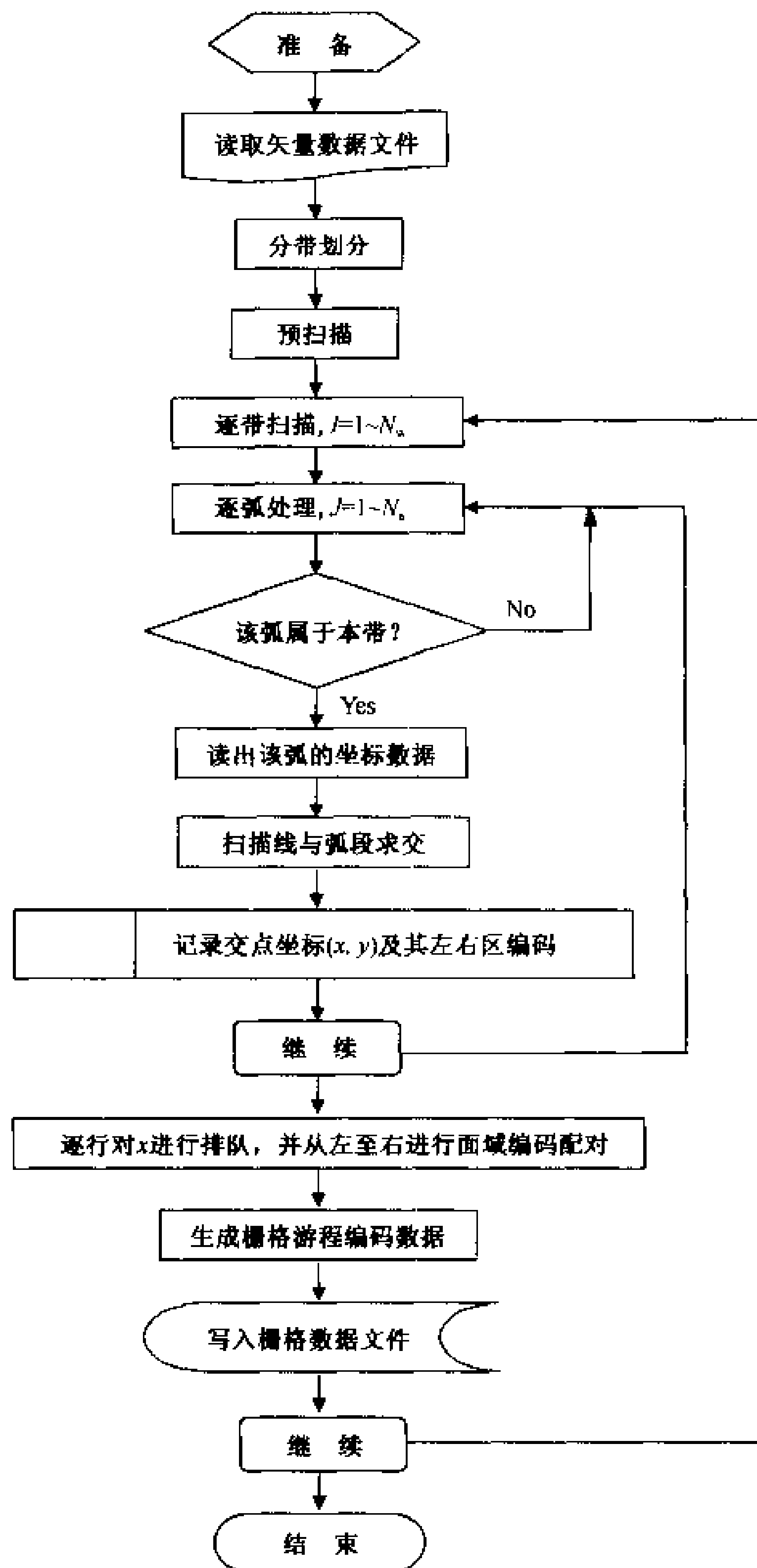


图 12-8 带扫描算法流程

12.1.4 复数积分算法

复数积分算法的实质是对整个栅格矩阵的栅格单元逐个判断其所属多边形的编号。判断方法是：由待判定点对每个多边形的封闭边界计算复数积分，若对于某多边形的积

分值为 $2\pi i$，则该待判定点属于此多边形，赋予该多边形编号；否则，该点位于此多边形外部，不属于该多边形。

复数积分算法涉及许多乘除运算，尽管可靠性好，设计也并不复杂，但运算时间很长，难以在较低档次的计算机上采用。若采用一些优化方法，如根据多边形边界坐标的最大最小值范围组成的矩形（也称多边形最小包容窗）来判断是否需要做复数积分运算，则可以在一定程度上提高运算速度。此处不赘述。

12.1.5 边界代数算法

边界代数算法是边界代数多边形充填算法（boundary algebra filling，BAF）的简称，也称炸弹算法。该法的实质是基于面域多边形模仿积分求面积的过程。如图 12-9 所示，以转换区域内仅有一个多边形为例，设多边形的编号为 a（可以据此对其属性进行索引），其原理如下：

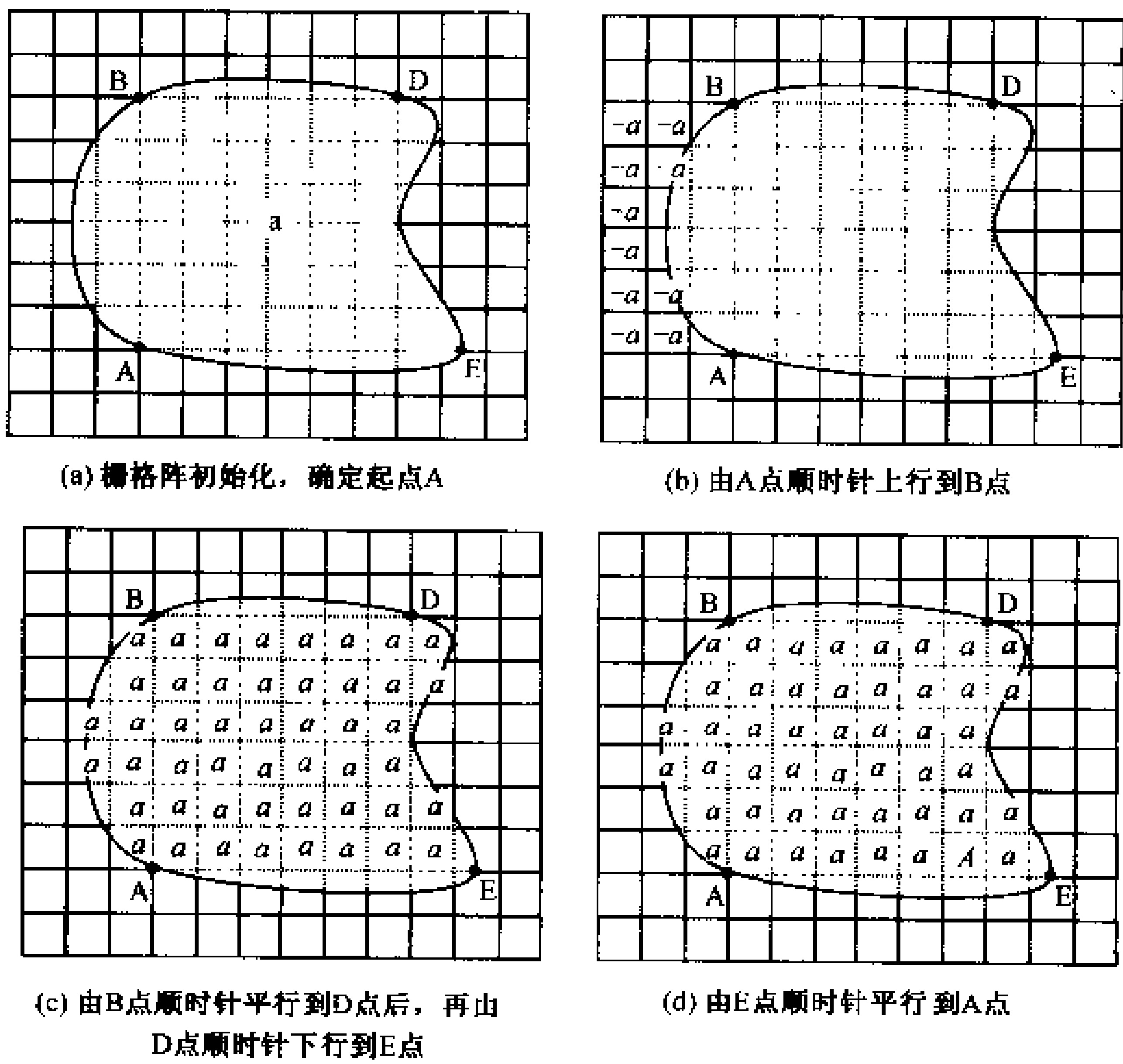


图 12-9　边界代数法基本原理

首先，将覆盖该多边形的面域(地图)进行整体栅格化，并对栅格阵进行零初始化。

然后，由其边界上某一点开始顺时针方向搜索其边界线，当边界线段为上行时，对该线段左侧具有相同行坐标的所有栅格全部减去一个 a；当边界线段为下行时，对该线段左侧（从前进方向看为右侧）具有相同行坐标的所有栅格全部加上一个 a；当边界线段平行于栅格行行走时，不做运算。

循环一周，回到起点，则该多边形边界内的栅格均被赋予了该多边形的编号 a，即多边形边界内的栅格均具有该多边形的属性，而多边形边界外的栅格值不变（仍为零）。到此，完成了该多边形由矢量向栅格的转换。

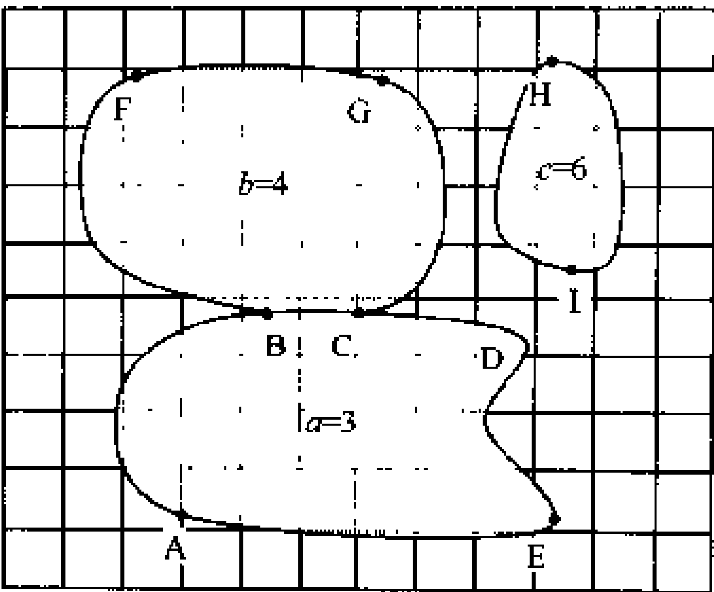


图 12-10　边界代数法实例

事实上，每幅地图都是由多个面域多边形组成的。如果将地图内不属于任何多边形的区域（包括无穷远点）看成一个编号为零的特殊区域，则每个多边形的每一条边界弧段均与两个不同编号的面域相邻，按其前进方向分别称为左、右多边形。据此，可以将上述基于单一多边形的边界代数法原理作以下扩展：

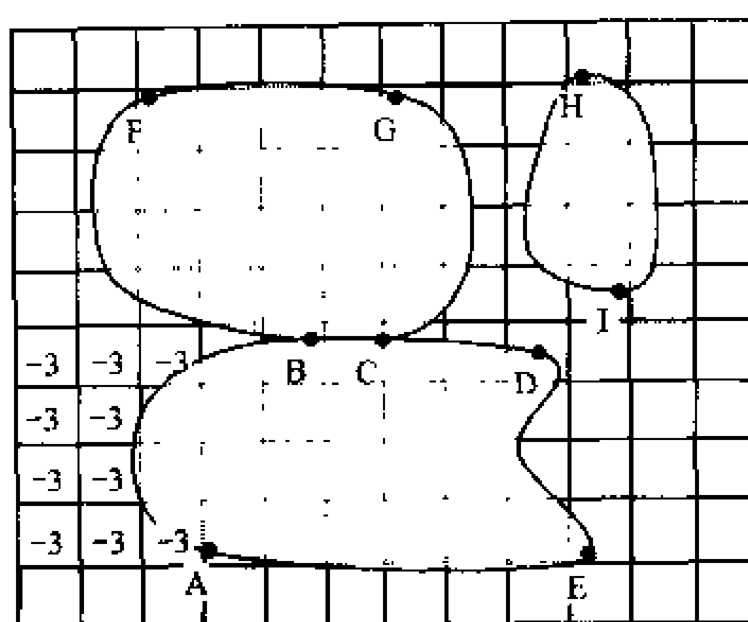
从某一位置开始，对图中的所有弧段逐条搜索，当弧段上行时，将该弧段与左图框之间的栅格加上一个值＝（该弧的左多边形编号）－（该弧的右多边形编号）；当弧段下行时，将该弧段与左图框之间内的栅格加上一个值＝（该弧的右多边形编号）－（该弧的左多边形编号）；当弧段平行栅格行行走时，不做运算。

为表述方便，以图 12-10 所示由 3 个多边形构成的图域为例，假定多边形 A 的属性编码 $a=3$；多边形 B 的属性编码 $b=4$；多边形 C 的属性编码 $c=6$；组成弧段为 AB、BC、CD、DE、EA、BF、FG、GC、CB（与 BC 为公共边）、HI 和 IH；按字母顺序为弧段前进方向，则可以分别确定上述弧段的行走性质及其左右多边形属性码；根据边界代数法原理，各弧与左图框之间内的栅格的代数运算值如表 12-3 所示：

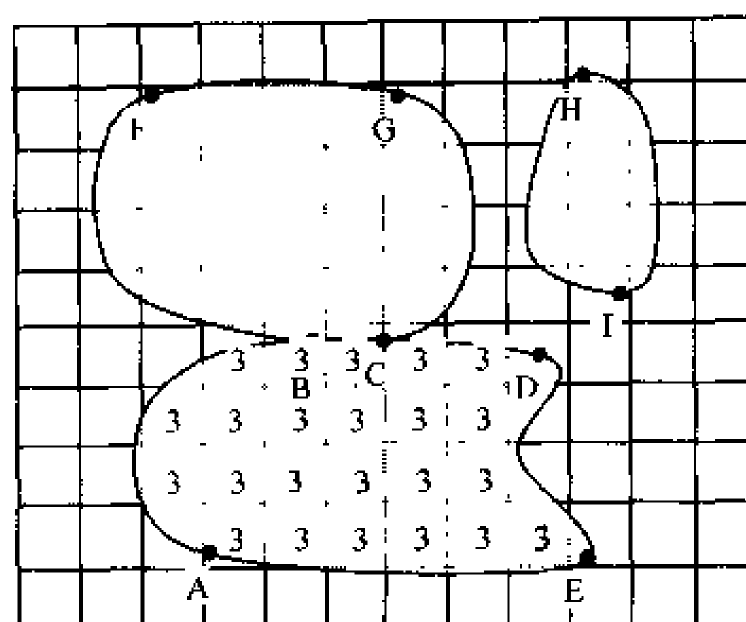
表 12-3　弧段行走性质及其左右多边形编码

弧段名称	行走性质	左多边形编码	右多边形编码	代数运算值
AB	上行	0	3	－3
BC	平行	4	3	0
CD	平行	0	3	0
DE	下行	0	3	3
EA	平行	0	3	0
BF	上行	0	4	4
FG	平行	0	4	0
GC	下行	0	4	4
HI	下行	0	6	6
IH	上行	0	6	6

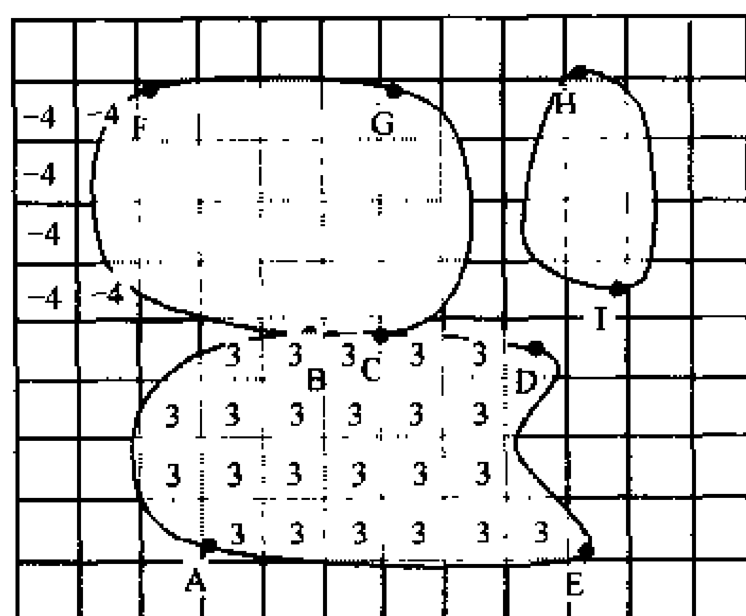
若按表 12-3 的顺序，依次对图 12-10 进行弧段扫描，其栅格化过程如图 12-11 所示。



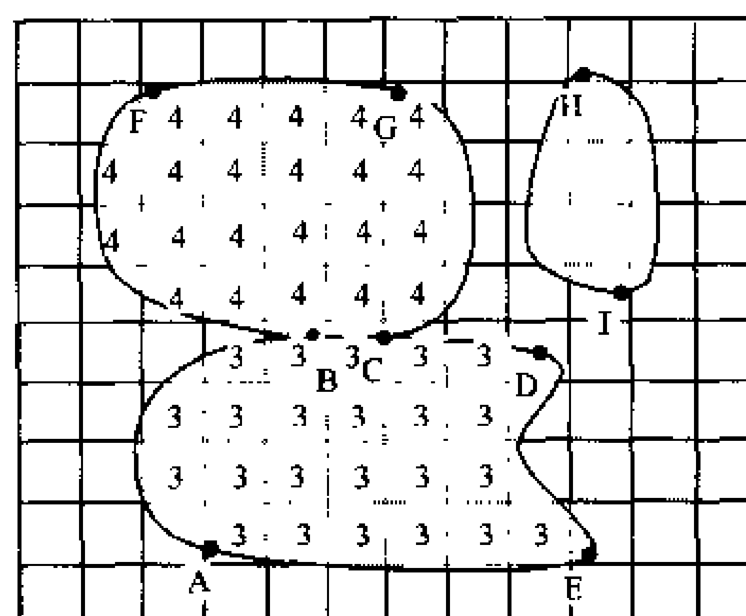
(a) 由A往B上行



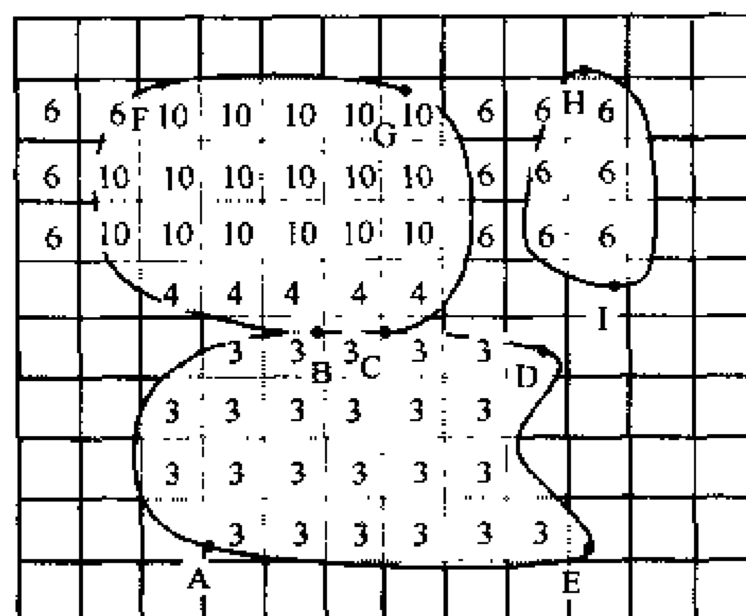
(b) 经BC、CD，由D往E下行，再到A



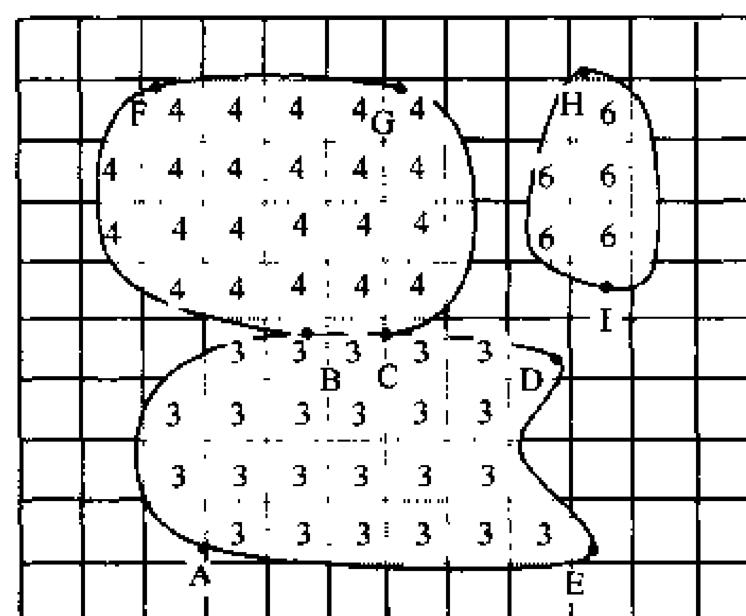
(c) 由B往F上行



(d) 从F经G往C下行，再经C到B



(e) 由H往I下行



(f) 由I往H上行，即到H

图 12-11　边界代数法运算过程实例

由此可见，扩展后边界代数法与其他算法的区别在于：它不是通过逐点搜寻来判断面域边界，而是根据矢量多边形边界的拓扑信息，通过简单的加减代数运算将拓扑信息赋予各栅格点。其特点是可以不考虑弧段的排列顺序，即无需考虑边界与搜索轨迹之间的关系，只需对每条弧段逐一搜索且仅一次，就可完成矢量向栅格的转换。因此，边界

代数法算法简单，可靠性好，运算速度快。

边界代数收敛法用数学语言描述为：

(1) 按矢量地图的最大、最小窗坐标 max (x, y) 和 min (x, y) 确定栅格区域范围；

(2) 选定合适的栅格尺寸，将栅格区域剖分为 N 行 M 列的矩阵 E；

(3) 将矩阵初始化，令 $E_{ij}=0$；

(4) 从某一多边形（设其属性编码为 a_k，$k=1\sim n$，n 为多边形总数）的某一边开始沿顺时针方向追踪，同时对所经过边的整体左侧的栅格属性值进行动态修改：

$$\left.\begin{array}{ll} E_{ij}=E_{ij}-a_k & y_i<y_{i+1}\text{ 时} \\ E_{ij}=E_{ij}+a_k & y_i>y_{i+1}\text{ 时} \\ E_{ij}=E_{ij} & y_i=y_{i+1}\text{ 时} \end{array}\right\} \tag{12-8}$$

(5) 判断是否所有多边形的所有弧段均搜索完毕；若非，则继续搜索；

(6) 若是，则完成搜索，并形成栅格矩阵的最终记录。

§12.2 栅格数据向矢量数据转换

栅格数据向矢量数据转换的目的有三：其一为数据入库，其二为数据压缩，其三为矢量制图。以数据压缩为例，目前一般扫描仪的扫描精度均可以达到 800DPI（800 dots/inch，相当于每个栅格的尺寸为 0.0125 mm）以上。以一条 0.1 mm 宽的线条而言，其扫描后横断面也占 8 个栅格。而按矢量数据的要求，一条线的宽度必须而且只能是一个栅格的宽度。数据入库工作，也主要指如何将自动扫描仪获取的栅格数据转入到矢量数据库中进行管理。

点的栅格数据向矢量数据转换，就是将栅格点的中心转换为矢量坐标的过程；弧段的栅格数据向矢量数据的转换，就是提取弧段栅格序列点中心的矢量坐标的过程；面域多边形的栅格数据向矢量数据转换，则是提取具有相同属性编码的栅格集合的矢量边界及边界与边界之间拓扑关系的过程。以栅格多边形的矢量化转换为例，其一般过程为：

(1) 多边形边界提取：采用高通滤波将栅格图像二值化或以特殊值标识边界点；

(2) 边界线搜索：逐弧进行，由某一节点开始沿某一方向进入，朝该点的 8 个邻域搜索其后续节点，直到连成弧段；

(3) 拓扑关系生成：对于已经用矢量表示的边界弧段，判断其与原图中各多边形的空间关系，形成完整拓扑结构并建立与属性数据的联系；

(4) 去除多余点并进行曲线圆滑：由于搜索是逐个栅格进行的，必然造成多余点记录，为减少数据冗余，必须去除。此外，由于栅格精度的限制，边界线搜索的结果曲线可能不够光滑，需要采用一定的插补算法进行光滑处理（详见 11.3 节）。

通常，根据栅格数据矢量化过程的不同，将它分为基于图像处理的矢量化和基于窗口匹配的矢量化两类。

12.2.1 基于图像处理的矢量化算法

基于图像处理的矢量化主要是对点和线状地物特征的提取。包括三个主要步骤：即二值化、细化和跟踪。

1. 二值化

所谓二值化，是在一个设定的灰度阈值的基础上，对扫描获得的灰度图像（如 256 级灰度）进行 0 或 1 的简化处理，即有无判断。算法如下：

$$B(i,j) = \begin{cases} 1 \cdots\cdots G(i,j) \geqslant T \\ 0 \cdots\cdots G(i,j) < T \end{cases} \tag{12-9}$$

式中：$B(i, j)$ 为第 i 行、j 列栅格的属性编码；$G(i, j)$ 为第 i 行、j 列栅格的灰度值；T 为灰度阈值。

对于彩色密度图像和遥感伪彩色图像，需要使用动态跟踪和动态阈值处理技术。限于篇幅，此处不再赘述，可参看有关遥感图像处理资料。

2. 细化

对二值化处理后的点或曲线进行细化的方法有多种，代表性的有剥皮法和骨架法。

1）剥皮法的原理

从线的边沿两侧开始，每次剥去等于一个栅格宽度的一层，直到最后仅剩下彼此相连的两个栅格宽或恰好一个栅格宽的线划图形为止。

当还剩两个栅格宽时，要根据栅格单元边相连或角相连的规则来判断是否能继续剥皮。此时可以使用 3×3 窗口匹配原理，即通过检查中心栅格与其 8 领域栅格的连通性，来判断该中心栅格是否可以被删除。如图 12-12 所示，其中 12-12（a）、（b）的中心栅格不可删除；而图 12-12（c）、（d）的中心栅格删除后仍能保持邻域栅格的角相连。

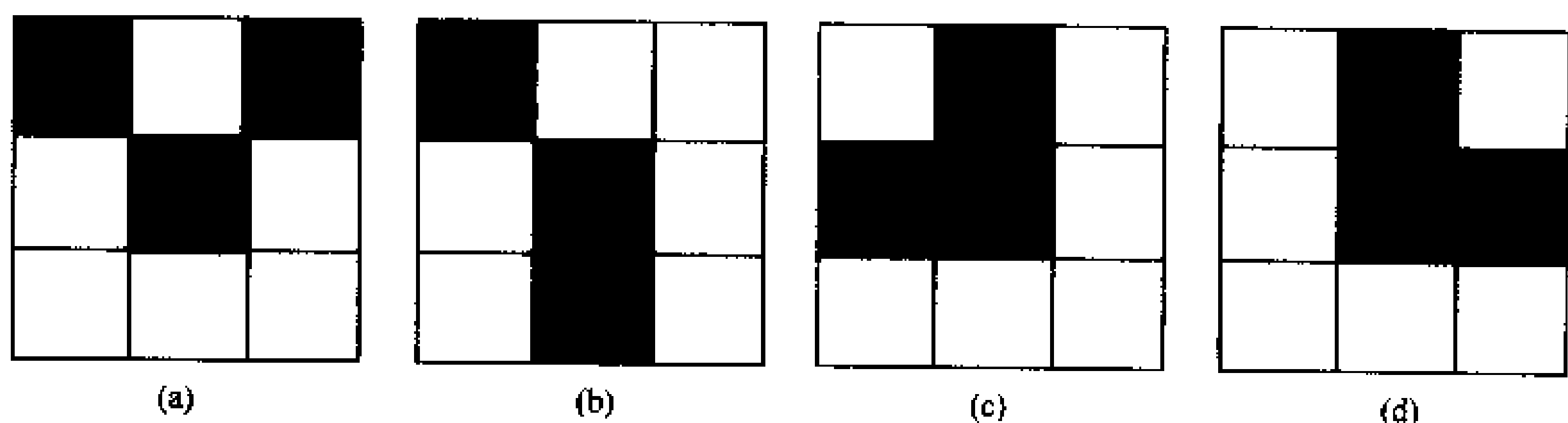


图 12-12　3×3 窗口匹配剥皮法原理

由于每个栅格有 0 和 1 两种可能，因此，每一个中心栅格的 3×3 窗口共有 2^8 种（256 种）可能模式。考虑到 90°、180°和 360°旋转，以及镜像、反射等重复模式，共剩下 51 种不同的模式。其中，还有 19 种可将中间点删除，读者可自行检测。

2）骨架法的原理

在二值化的基础上，针对各条待矢量化的栅格线，求出线上每一栅格的 3×3 窗口的属性码之和，并用此和值对该中心栅格重复赋值；对重新赋值后的栅格线找到每一行中最大栅格属性码所在位置，确定该栅格位置即为栅格线的骨架，即是最后的矢量线所经过的点。

以图 12-13 为例，展示了骨架法的实施过程和结果。

1	1	0	0	0	0
1	1	1	0	0	0
0	1	1	1	0	0
0	0	1	1	0	0
0	1	1	1	0	0
0	0	1	1	0	0

(a) 二值化结果

4	5	3	1	0	0
5	7	6	3	1	0
3	6	7	5	2	0
2	5	8	6	3	0
1	4	7	6	3	0
1	3	5	4	2	0

(b) 3×3窗口属性码求和

0	1	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0

(c) 骨架法提取结果

图 12-13 骨架法原理及实施实例

3. 跟踪

所谓跟踪，是指将细化后的栅格数据整理为从结点出发的弧段或封闭曲线，并以矢量的形式存储特征栅格点的坐标。其实施过程如下：

（1）从图幅西北角开始，用水平扫描线检测线段的起点，并以该起点为当前点，置数组或缓冲区为空；

（2）以当前点为中心，按逆时针方向检测其 8 邻域，寻找属性编码相同的下一点；

（3）若找到，则将当前点的坐标放入数组或缓冲区；并将所找到的点置为当前点，转第（2）步；

（4）若未找到，则跟踪结束，将当前点的坐标放入数组或缓冲区；

（5）将数组或缓冲区中的全部坐标数据顺序写入弧段数据文件，并将数组或缓冲区重新置空；

（6）判断是否跟踪完毕：若否，则转第（1）步；若是，则关闭数据文件，完成第一条弧段或封闭曲线的矢量化。

基于图像的矢量化技术的难点是，对各类注记等复杂图形的提取和识别，以及对断裂等非连续现象的识别和自动处理。

12.2.2 基于窗口匹配的矢量化算法

基于窗口匹配的矢量化核心是：多边形边界线搜索采用 2×2 的栅格窗口，在每个窗口内的四个栅格的数据模式可以惟一地确定下一弧段的搜索方向和该弧段的拓扑关

系。这一方法可以加快搜索速度，拓扑关系也容易建立。具体步骤如下：

(1) 边界点和节点提取：采用2×2的栅格窗口阵列作为搜索窗口，顺序沿行、列方向对栅格全图进行扫描。如果窗口内4个栅格有两个以上的属性编码，则该4个栅格为边界点并保留各栅格的原有属性编码；若窗口内4个栅格有3个以上不同属性编码，则标识为节点（即不同边界弧段的图形不连通），并保留各栅格的原有属性编码；若对角线上栅格属性编码两两相同，也属于不连通情况，作为节点处理。边界点有6种结构，如图12-14所示；而节点有8种结构，如12-15所示。

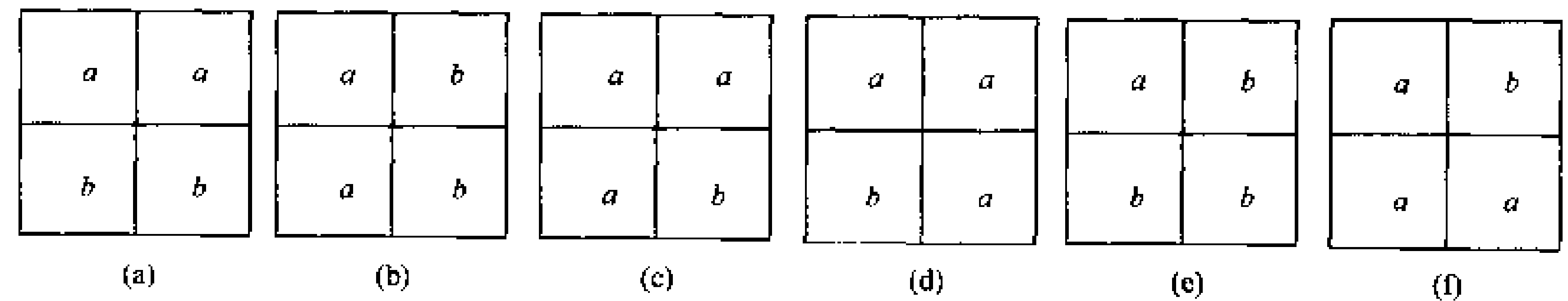


图12-14 边界点的6种结构

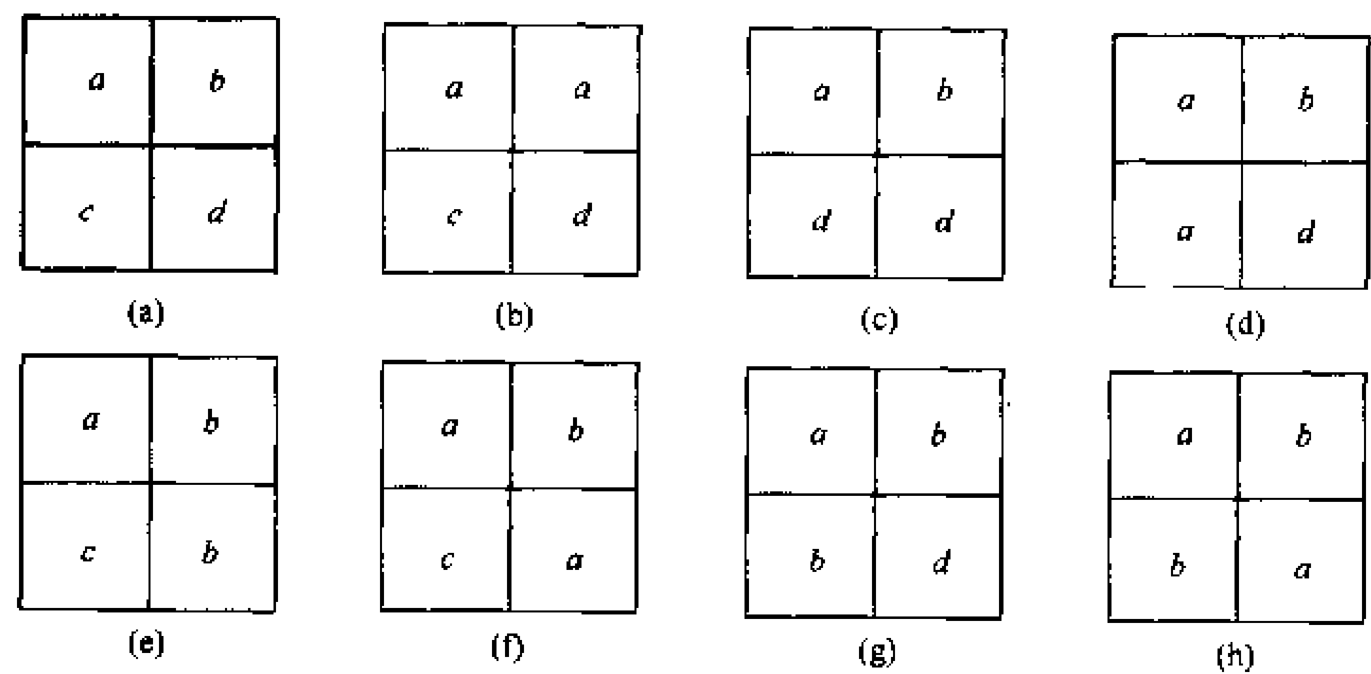


图12-15 节点的8种结构

(2) 边界搜索与左右多边形信息记录：逐条进行弧段搜索。每条弧段从一组已标识的四邻节点开始，首先记录起始边界点的两个多边形编号作为该弧段的左右多边形；下一节点的搜索方向则由前一节点的进入方向和该点的可能走向来决定。每个节点只有两个可能的走向：或者是前一节点的进入方向，或者是后一节点的可能进入方向，如图12-16所示。

栅格数据向矢量数据转换的关键是边界线搜索、拓扑结构生成和多余点处理。任伏虎（1991）基于四窗口匹配技术，发展了一种双边界直接搜索法（double boundary direct finding, DBDF），较好地解决了上述问题。该法的基本思想是：通过边界提取，将边界的左右多边形信息保存在边界点上，每条边界弧段由两个并行的边界链组成，分别记录该边界弧段的左右多边形编号。

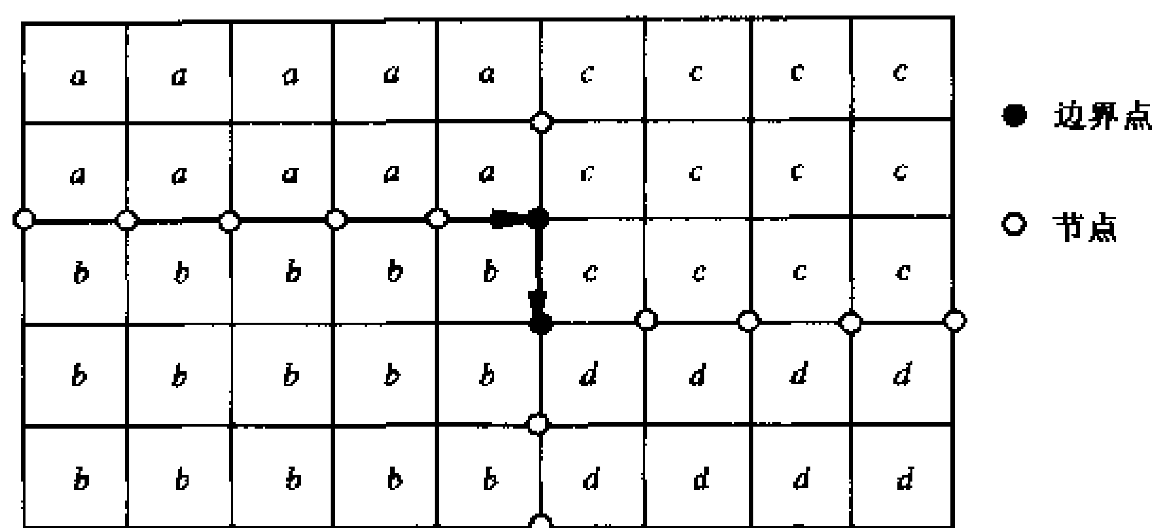


图 12-16 边界搜索进程

§12.3 TIN 向规则格网 DEM 转换

数字地面模型（digital terrain model，DTM）或称数字地形模型，被定义为以数字形式存储的地球表面所有信息的总和。它是 MIT 的 C.L.Miller 在解决道路设计问题时，于 20 世纪 50 年代中期提出的概念。相关概念还有：数字高程模型（digital elevation model，DEM），被定义为以数字形式存储的表示物体位置高程值的集合，是用数字形式描述的地形表面；数字地貌模型（digital geomorphic model，DGM），被定义为描述地表起伏的某类数据的集合或其总称。许多场合下，往往将 DTM、DEM 等同看待。20 世纪 60 年代以来，数字高程模型的研究和应用得到迅猛发展（王家耀，2001），研究内容涵盖数据采集、精度分析、粗差探测、质量控制、地形分类、数据压缩、TIN 构建、工程应用、军事应用、地形表面分析等，并且涌现了一批优秀的 DTM 软件包，如德国 Stuttgart 大学的 SCOP，Munich 工业大学的 HIFI，Hannover 大学的 TASH，奥地利 Vienna 工业大学的 SORA，以及瑞士 Zurich 大学的 CIP 等。DTM 已经成为 GIS 理论、方法与技术的一个重要组成部分，空间数据的 2.5 D 或 2.75 D 显示、地形分析等都是以 DTM 为基础的。

此外，DTM 还是 GIS 数据库的基本内核，通常采用 4 种数据结构来表达 DTM，即不规则三角网（triangular irregular network，TIN）、等高线图（contour line map，CLM）、断面线（section lines，SLs）、规则格网（regular grids，RGs）及其混合（如 Grid + TIN）。其中 TIN 直接采用离散采样点构网，适合任意复杂地形模拟，而且易表达拓扑关系、任意点内插方便、内插速度快。而基于 RGs 的 DEM 虽然通过原始数据插值得到，但也具有易表达拓扑关系、任意点内插方便、内插速度快等特点，且便于地形分析和计算，适合计算机数据处理。为发挥两者的特点，讨论由 TIN 内插生成 DEM 很有必要。

从 TIN 内插生成格网 DEM 有多种算法。但大多数算法均将三角形视为平面，按线性内插方式计算出落在每一三角形中的格网点的高程值。这必然导致内插后的格网 DEM 质量不高，具体表现为在三角形公共边处（尤其是小夹角时）容易出现异常的地形棱角。为了消除这些棱角，Preusser（1984）曾提出了 C 连续双 5 次多项式插值方法。

但是，该法算法复杂，不便应用。王家耀（2001）则基于所定义的磨光函数，提出了一种在两片三角形公共边附近的一定区域内，由 TIN 到规则格网 DEM 光滑内插的算法。

1. 磨光函数与磨光宽度

磨光函数的基本思想是通过对一个函数积分后再取差商，以得到一个与原函数近似的函数。例如，函数$f(x)$的 h 步长的对称差商为：

$$\begin{aligned}\frac{\delta_h f(x)}{h} &= \frac{1}{h}\left[f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)\right] \\ &= \frac{1}{2h}\left[\left(x+\frac{h}{2}\right)_+^2-\left(x-\frac{h}{2}\right)_+^2\right] \\ &= \begin{cases} x & x > h/2 \\ 1/2h(x+h/2)^2 & -h/2 \leqslant x \leqslant h/2 \\ 0 & x < -h/2 \end{cases}\end{aligned} \tag{12-10}$$

显然，式(12-10)与 $y=x$，很相似，仅在 $-h/2\leqslant x\leqslant h/2$ 这一段上有差异：

$$\frac{1}{2h}\left(x+\frac{h}{2}\right)^2-x_+$$

这一差异是与 h 同阶的量，且在 $x=-h/2$ 和 $x=h/2$ 处分别切于原折线 $y=x_+$，两者的图形比较见图 12-17。

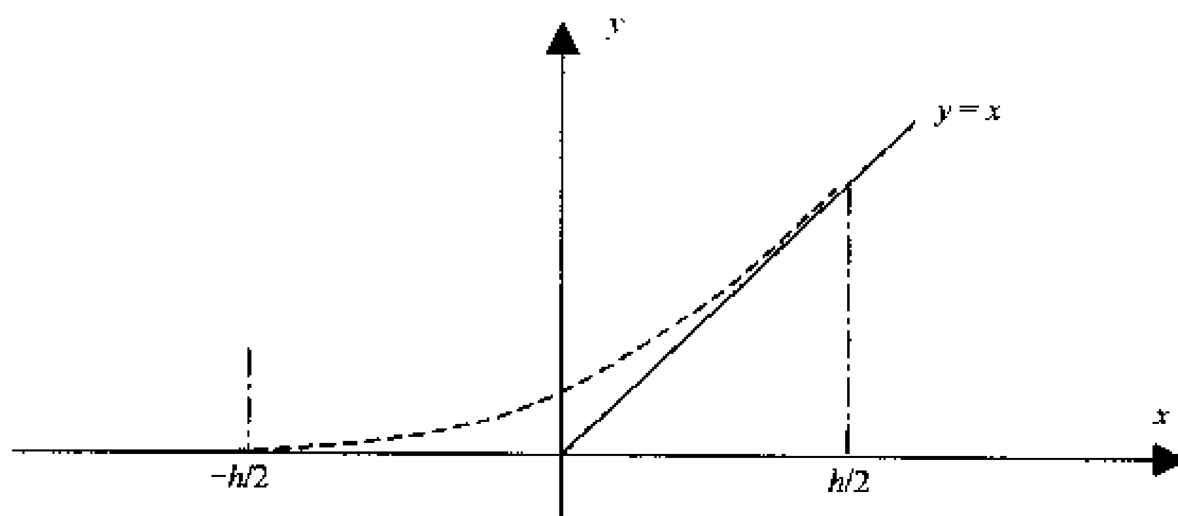


图 12-17　差商函数原理

这种先积分一次再实行差商所得的函数称为磨光函数(王家耀，2001)，它和原函数有较好的近似性质，且提高了光滑性。其一般性定义为：

设有函数 $f(x)$，对函数 $f(x)$作一次不定积分后，再作一次以 h 为步长的对称差商，所得的新函数

$$f_1(x) = \frac{\delta_h}{h}\int f(t)\mathrm{d}t = \frac{1}{h}\int_{x-h/2}^{x+h/2} f(t)\mathrm{d}t \tag{12-11}$$

为原函数 $f(x)$的一次磨光函数，即对函数 $f(x)$进行了一次磨光，h 称为磨光度。

2. 两直线的磨光函数

如图 12-18 所示，折线 L 的方程为：

$$L: y = \begin{cases} a_1 x, \cdots, x > 0 \\ a_2 x, \cdots, x < 0 \end{cases} \tag{12-12}$$

若采用式(12-11)对其磨光，经积分运算求得其磨光函数为：

$$f_1(x) = \begin{cases} a_1 x \quad x > h/2 \\ \dfrac{x^2}{2h}(a_1 - a_2) + \dfrac{x}{2}(a_1 + a_2) + \dfrac{h}{8}(a_1 - a_2) \quad -h/2 \leqslant x \leqslant h/2 \\ a_2 x \quad x < -h/2 \end{cases} \tag{12-13}$$

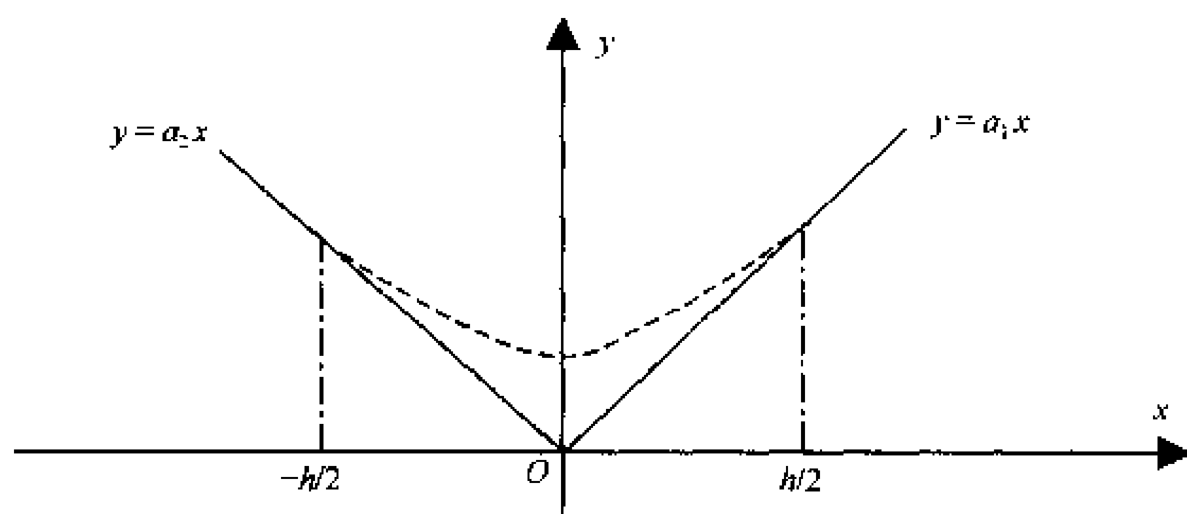


图 12-18　折线磨光

3. 两三角形接边处磨光

如图 12-19 所示，△ABC 和△ABD 为两个以 AB 为公共边的空间三角形，DD' 和 CC' 分别垂直于 AB，D'、C' 为垂足。设 A、B、C、D 的坐标分别为(x_a, y_a, z_a)、(x_b, y_b, z_b)、(x_c, y_c, z_c)和(x_d, y_d, z_d)。该两三角形内插为规则格网后的磨光过程包括空间变换、直线磨光和逆变换。

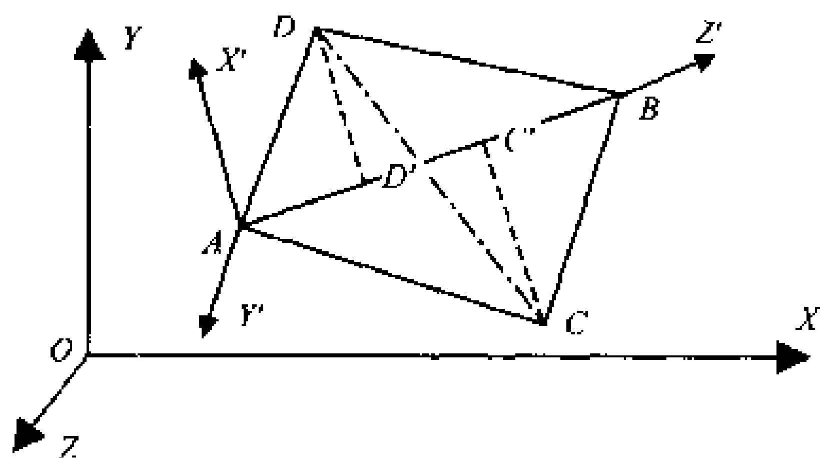


图 12-19　三角形磨光坐标平移与旋转

1) 空间变换

为便于计算和分析，将上述三角形作空间变化，使 AB 平行于 Z 轴，且 A 与原点重合。因此，需要经过平移和旋转变换。

首先平移：按下式将坐标原点移至 A 点：

$$\begin{cases} x' = x - x_a \\ y' = y - y_a \\ z' = z - z_a \end{cases} \tag{12-14}$$

然后旋转：以 A 为原点，AB 为 z' 轴，x' 轴位于 $\triangle ABD$ 内，y' 轴垂直于 $\triangle ABD$。

设原始坐标系为 $O\text{-}XYZ$；平移旋转后的坐标系为 $A\text{-}X'Y'Z'$。令 $A\text{-}X'Y'Z'$ 中 X'、Y'、Z' 轴关于 $O\text{-}XYZ$ 的方向角如表 12-4 所示。

表 12-4　坐标平移旋转后的方向角

	Ox	Oy	Oz
Ox'	α_1	β_1	γ_1
Oy'	α_2	β_2	γ_2
Oz'	α_3	β_3	γ_3

设任意点 P 在 $O\text{-}XYZ$ 和 $A\text{-}X'Y'Z'$ 中的坐标分别为 (x, y, z) 和 (x', y', z')，则它们的变换公式为：

正变换：

$$\begin{cases} x' = x\cos\alpha_1 + y\cos\beta_1 + z\cos\gamma_1 - x_a \\ y' = x\cos\alpha_2 + y\cos\beta_2 + z\cos\gamma_2 - y_a \\ z' = x\cos\alpha_3 + y\cos\beta_3 + z\cos\gamma_3 - z_a \end{cases} \tag{12-15}$$

逆变换：

$$\begin{cases} x = x'\cos\alpha_1 + y'\cos\beta_1 + z'\cos\gamma_1 + x_a \\ y = x'\cos\alpha_2 + y'\cos\beta_2 + z'\cos\gamma_2 + y_a \\ z = x'\cos\alpha_3 + y'\cos\beta_3 + z'\cos\gamma_3 + z_a \end{cases} \tag{12-16}$$

2）直线磨光

显然，$\triangle ABC$ 和 $\triangle ABD$ 在 $X'AY'$ 面上的投影均为一条交于原点 A 的直线。设 $\triangle ABC$ 和 $\triangle ABD$ 的投影线关于 X' 轴的斜率分别为 a_1 和 a_2。可知，$a_1 = 0$，如图 12-20 所示。

为适用式(12-13)的磨光函数，图 12-20 须与图 12-18 对应。令 $X'AY'$ 饶 Z' 轴旋转一个角度 θ，使 Y'' 恰好位于两条投影线的角平分线上，如图 12-21 所示。

设任意点 P 在 $X'AY'$ 和 $X''AY''$ 中的坐标分别为 (x', y', z') 和 (x'', y'', z'')，则它们的变换公式为：

正变换：

$$\begin{cases} x'' = y'\sin\theta + x'\cos\theta \\ y'' = y'\cos\theta - x'\sin\theta \\ z'' = z' \end{cases} \tag{12-17}$$

逆变换：

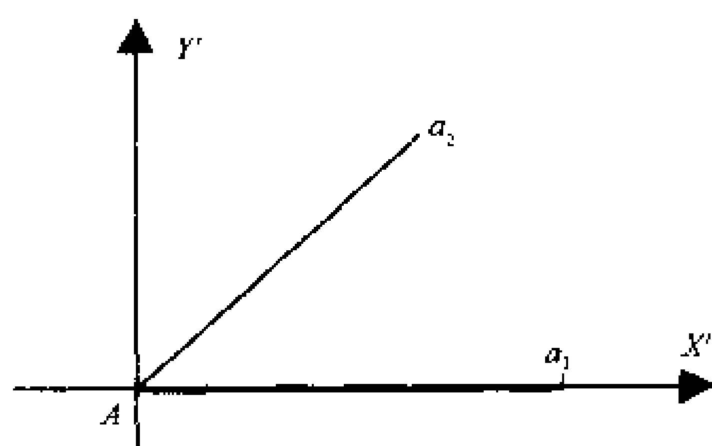


图 12-20 三角形投影线及其斜率

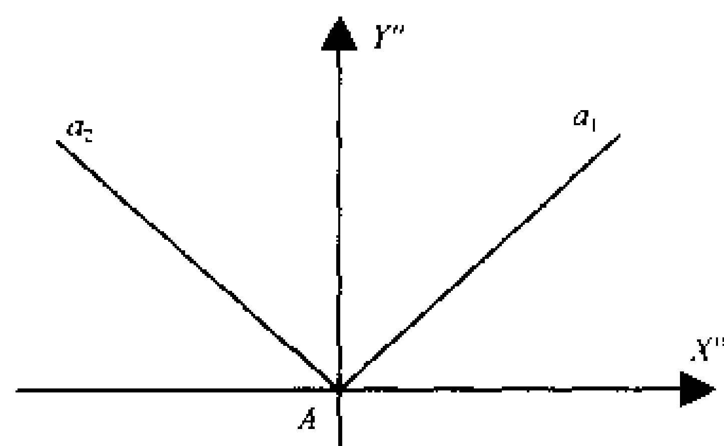


图 12-21 三角形投影线的旋转处理

$$\begin{cases} x' = x''\cos\theta - y''\sin\theta \\ y = x''\sin\theta + y'\cos\theta \\ z' = z'' \end{cases} \tag{12-18}$$

于是两条投影线的斜率 $a_1 = -a_2$，式(12-13)的磨光函数可简化为：

$$f_1(x) = \begin{cases} a_1 x & x > h/2 \\ \frac{x^2}{h}|a_1| + \frac{h}{4}|a_1| & -h/2 \leqslant x \leqslant h/2 \\ a_2 x & x < -h/2 \end{cases} \tag{12-19}$$

选取合适的磨光宽度 h，即可以对两条投影线进行磨光。

3）逆变换

首先按平面方式内插出规则格网点，然后根据公式（12-15）、（12-17）和（12-19），可以对两个三角形上的所有格网点进行磨光处理。之后，根据公式（12-18）、（12-16）进行逆变换回到原始坐标系，即可以得到一个光滑的格网曲面。

4．磨光内插步骤和算法分析

由 TIN 到格网 DEM 的磨光内插的具体步骤如下：

（1）根据 TIN 的拓扑表，找出所有具有公共边的三角形，并按顺序存入一个文件；

（2）将所找出的每个三角形按其重心分割为 3 个子三角形，找出具有公共边的所有子三角形对偶；

（3）确定格网 DEM 的纵横方向的步长和磨光宽度；

（4）按格网 DEM 的纵横方向的步长，求出落在各子三角形中的所有格网点，并按线性内插原理求出各格网点的 z 值；

（5）按公式（12-15）、（12-17）和（12-19），在子三角形对偶中进行磨光计算；

（6）按公式（12-18）和（12-16），对磨光计算结果进行逆变换，即完成 TIN 数据到规则 DEM 数据的转换和磨光内插。

使用以上方法进行 TIN 数据到规则 DEM 数据的转换和磨光内插的关键，是选择合适的磨光宽度。由于磨光宽度很难用一定的规律进行控制，所以在磨光计算时，为得到较好的磨光效果，且保留 TIN 的基本地形特征，可以针对不同的区域和地形特征设置

不同的磨光宽度。比如可令：当两个三角形的夹角小于某一角度时，磨光宽度等于预设宽度的两倍。此外，当格网点落在三角形顶点时，可以分别在两组以上的子三角形对偶中进行磨光内插，导致高程结果不一致，对此简单的处理措施是取其平均值。

参 考 文 献

龚健雅. 2001. 地理信息系统基础. 北京：科学出版社. 174～189

任伏虎. 1991. 遥感与GIS集成软件系统GRAMS. 北京：中国首届青年GIS与遥感学术讨论会论文

王家耀. 2001. 空间信息系统原理. 北京：科学出版社. 197～203

Miller C. L. and Laflamme R. A., 1958. The Digital Terrain Model: Theory and Application. Photogrammetric Engineering 24. 433～422

Weibel R. and Heller M., 1991. Digital Terrain Model. In: *GISs: Principles and Applications*. Eds. Maguire D J, Goodchild M F, and Rhind D W. Vol. 1. 269～297

Preusser A., 1984. Bivariate Interpolation ueber Dreieckselementen durch Polynome 5. Ordung mit C1-Kontinuitaet. Zeitschrift fuer Vermessungswesen. 109, Heft 6. 292～301

Ebner H and Tang L., 1989. High Fidelity Digital Terrain Models from Digitized Contours. 14th ICA Congress. Budapest, Hungary.

第十三章　空间数据误差分析算法

数据采集是空间信息系统建设项目中最昂贵的部分，Blakeman（1987）、Morse 和 Hovey（1990）的研究结果表明，GIS 中数据采集的费用通常占总项目费用的 80%。所以近年来，GIS 数据质量问题及由此产生的 GIS 数据质量控制技术成为了国内外学术界研究的热点问题之一。

对 GIS 来说，空间数据是基础；非空间数据是内涵，是地理单元的纵深描述。随着 GIS 产业的兴起和发展，GIS 的建立及其功能的改善对 GIS 空间数据的质量等基础理论问题提出了更高的要求。GIS 的数据质量主要是指空间数据的质量，GIS 的空间数据主要来源于直接数字测量系统、遥感与摄影测量系统、局域和广域差分 GPS 系统、手工数字化和扫描数字化系统等。GIS 空间数据的不确定性可以看做是一种广义的误差，它包括可度量和不可度量的误差，也包括数值和概念上的误差。空间数据的随机性、数据与概念的复杂性、模糊性和不完整性等，都可以视作不确定性问题所包含的内容。因此，GIS 空间数据的质量和质量控制研究与不确定性研究密切相关；既需要应用经典的误差分析技术与测量平差理论，也需要应用各种新的理论来研究不确定性的表达、度量及质量控制方法。

§13.1　空间数据误差分类与来源

空间数据误差是无法回避的事实存在。在复杂多变的现实世界中，空间实体多相互混杂，空间数据多种多样，大量复杂的地理现象定义模糊，界限有时不很分明，也并非都是空间匀质分布的。使用空间数据描述空间实体，不可避免地要给出空间实体的定义（史文中，1998），并对复杂的空间实体进行取舍、抽象和离散（Zhang，1996）。在定位、采集、录入、储存、处理、分析和管理表达空间实体的空间数据时，常常不易明确地将空间实体分类，不得不从中抽取重要的特性来表达确定或不确定的现实实体（如目标模型的空间点、线、面抽象），用确定的模型或规则描述空间实体，但是纯几何意义上的点、线和面在现实世界中并不存在，故数据采样只是一种近似描述，所获信息不可能穷尽现实世界的全部，其准确性也同时会受到传感器和数学模型等设备技术因素的影响。而且，获取的信息在被导入计算机系统并用于空间分析决策的过程中，又被部分舍弃或删除（如制图综合）。同时，获取大量空间数据的真值并不容易，甚至有些空间数据的严格或绝对意义上的真值往往并不存在。所以，空间信息系统中实体的位置和属性信息与现实实体存在差异，即空间数据常常含有误差，并在数据流中传播。

1. 空间数据误差分类

空间数据误差一般指测量值或计算值与真实数据或假定真实数据间的差值（Mikhail，Ackermann，1976）。这种误差可分为明显误差、源于自然或原始量测值的误

差和源于数据处理的误差（Burrough，1999）。根据误差的性质，可以分为系统误差、随机误差和粗差。根据误差的内容，则可分为位置误差、属性误差、完整性误差、逻辑一致性误差和时间误差等。位置误差表示测定位置和实际位置的差异，与位置坐标及其相应的数据类型有关，可以用位置准确度、几何准确度和拓扑准确度等定义。属性误差用以衡量属性数据量测值和其真值的差异，有定量和定性两种精度，其描述受到数据处理方法、分类算法、位置精度、属性特征的时间变化、卫星影像分辨率和光谱特性等很多因素的影响。逻辑一致性指计算机化了的空间实体满足空间实体间拓扑一致性、数据结构内部一致性、数据规范逻辑一致性的程度。完整性指数据库中因省略某些数据而带来的误差，定义了概括和抽象的程度。时间误差指空间数据库满足最新性需要的程度。

2. 空间数据误差来源

空间数据误差的来源很多，且与所使用的数据、算法和数据结构皆有关，主要来自属性的定义、数据源、数据建模和分析过程中引入的误差等。其中，数据源的误差又源于数据采集过程中的量测、人为判断和假设等（刘文宝等，2000）。对 GIS 空间矢量数据不确定性的进一步讨论主要围绕空间数据位置误差的分布、不确定性模型以及平差纠正等方面（刘大杰、史文中等，1999）。利用遥感技术获取的土地覆盖分类数据的空间数据误差，则主要来自空间、光谱、波谱和时域特性的误差（史文中，1998）。对于属性精度的研究一方面围绕属性特征分类精度的描述和精度指标的确定，另一方面则以缺陷的形式描述和记录属性误差，并通过抽样统计的方法获得这一精度的描述（刘春，2000）。

§13.2 属性误差的分析算法

13.2.1 误 差 矩 阵

在 20 世纪 80 年代初，表达分类、信息提取精度的方式是用一个被称作非点位（Non-site-specific）精度的数字来表示。当从遥感影像数据上获取分类的属性数据以后，为估计这种分类精度，需要知道从遥感影像上获得的数据与参考标准数据的对比的偏差。这一结果显然与抽取的样本与参考数据的精度有关。作为这种分类属性误差的估计，误差矩阵可以描述抽样对比的结果。所谓误差矩阵就是一个维数为抽样量的方阵。矩阵的行代表属性的一个分类，而代表参考数据。一般情况下，某一属性的分类应该是正确的，然而经常会发生误分的情况，而该类属性误分将被记录在误差矩阵中。

误差矩阵可以有效地记录属性数据分类的误差情况。表 13-1 给出了一个误差矩阵的例子。

总体误差 $=(65+81+85+90)/434=321/434=74\%$

生产误差：

$$A=65/75=87\%$$
$$B=81/103=79\%$$
$$C=85/115=74\%$$

$$D = 90/141 = 64\%$$

用户误差：

$$A = 65/115 = 57\%$$
$$B = 81/100 = 81\%$$
$$C = 85/115 = 74\%$$
$$D = 90/104 = 87\%$$

表 13-1 误差矩阵的例子

数据		参考数据(j)				
		A	B	C	D	行统计
分类数据(i)	A	65	4	22	24	115
	B	6	81	5	8	100
	C	0	11	85	19	115
	D	4	7	3	90	104
	列统计	75	103	115	141	434

从误差矩阵中获得总体误差的同时，也可以得到生产误差和用户误差两个描述性指标来描述属性分类的误差。这三类误差综合起来可以作为属性分类误差精度的一个指标。从表 13-1 所示的误差矩阵中，可以得到属性分类的总体误差为 74%。然而对于各种分类属性的精度，也需要一个详细的误差指标描述，所以对于生产误差，不同的分类属性可得到不同的精度。同样，也可得到不同分类属性的用户误差精度指标。

从数学描述上，假设有 n 个抽样样本分布在 k^2 欧氏空间里，每一个样本数据从遥感数据分类中，都可以确定为属于 k 类属性中的一类。同样这 n 个抽样样本都有相应的参考属性分类。把抽样对照的结果置于表 13-2，令 n_{ij} 表示获得的应该属于第 i 类的数据却分类为第 j 类属性数据的样本数量（$i=1, 2, \cdots, k$，$j=1, 2, \cdots, k$）。

表 13-2 误差矩阵的数学描述

数据		参考数据(j)				
		1	2	…	k	行统计 n_{i+}
分类数据(i)	1	n_{11}	n_{12}	…	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	…	n_{2k}	n_{2+}
	…	…	…	…	…	…
	k	n_{k1}	n_{k2}	…	n_{kk}	n_{k+}
	列统计 n_{+j}	n_{+1}	n_{+2}	…	n_{+k}	n

属性分类入第 i 类的样本数据总数为

$$n_{i+} = \sum_{j=1}^{k} n_{ij} \tag{13-1}$$

根据参考数据，属性分类应属于第 j 类的样本数据总数为

$$n_{+j} = \sum_{i=1}^{k} n_{ij} \tag{13-2}$$

遥感属性分类数据与参考数据对比，整体误差可以为

$$\text{Overall} = \frac{\sum_{i=1}^{k} n_{ii}}{n} \tag{13-3}$$

生产误差可以为

$$\text{Producer}_j = \frac{n_{jj}}{n_{+j}} \tag{13-4}$$

用户误差可以为

$$\text{User}_i = \frac{n_{ii}}{n_{i+}} \tag{13-5}$$

在误差矩阵的基础上，除了通过上述的总体误差、生产误差和用户误差之外，还可以得到其他一些描述分类误差的指标，所以矩阵误差是对属性误差分类描述的前提，具有非常的意义。在所有描述属性分类误差的指标中，Kappa 系数是应用最为广泛的一个指标。该指标是建立在统计意义上的，基于误差矩阵的一个描述分类误差的指标。进行 Kappa 分析的运算是 KHAT 统计，最早由 Cohen (1960) 提出，后来被广泛应用于遥感影像分类的精度估计中。KHAT 统计与其方差的数学表达为：

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}x_{+i}} \tag{13-6}$$

其中，+表示行或列的求和，r 为分类类别数，N 为总样本数，因此，KAPPA 系数的计算使用了误差矩阵中每一个元素。

KAPPA 值的意义可解释为：设某分类结果的 KAPPA 值为 0.7，则表示该分类方法以 70% 的程度优于随机地给像素赋予类别。

13.2.2 属性数据精度的缺陷率度量算法

所谓缺陷率简单地说就是单位数量数据单元中包含的缺陷数，由于这一指标计算中的缺陷数是属性数据质量的检验结果，所以通过有效质量检验结果获得的缺陷率指标值，可以用来度量属性数据的质量。

每一属性数据，对所检验样本的质量可采用计数缺陷的形式记录检验值，而缺陷类型已被现有国家测绘数据质量检验标准采用，即一条属性数据单元的缺陷分为三种类型，即一般缺陷（light _ def）、重缺陷（heavy _ def）和严重缺陷（serious _ def）。缺陷类型的划分可按现有质量检验标准执行。

对于一个待检验数据总体，有一个数据总体容量，可用 N 表示，它是指该检验数据的总体数据量，即如果检验过程中，以一个数据单元为检验对象，则总体容量就是整个检验总体中数据单元的总和。同样检验过程中有一个抽样容量，可用 n 表示，它是

指提供被检验的总抽取数据量，即如果检验过程中，以一个数据单元为检验对象，则抽样容量就是整个抽取数据单元的总和。

设有某一属性数据集合为 X，它包含数据单元 N 个（即总体容量为 N），采用不回放的简单抽样方法对数据单元进行抽样检查，对该集合抽样 m 次（或对 m 个性质相同的属性集合分别抽样），若每次抽样的样本容量为 n，抽样检验所得到的缺陷数为 y_i（$i=1, 2, \cdots, m$），有：

$$y_i = W_{SD_a}\cdot(\text{light_def}) + W_{SD_b}\cdot(\text{heavy_def}) + W_{SD_c}\cdot(\text{serious_def}) \tag{13-7}$$

其中，W_{SD_a}：一般缺陷在缺陷总和中的权重；W_{SD_b}：重缺陷在缺陷总和中的权重；W_{SD_c}：严重缺陷在缺陷总和中的权重。

根据国家测绘数据质量检验标准，从大量的实验数据分析认为，可以取 $W_{SD_a}=1$，$W_{SD_b}=3$，$W_{SD_c}=5$，当然这一取值可根据实际应用情况修正。

因每次抽样的样本容量为 n，所以对于其中第 i 次抽样有：

$$y_{ij} = \begin{cases} 1, & \text{若第 } j \text{ 个数据单元是有缺陷的} \\ 0, & \text{若第 } j \text{ 个数据单元是无缺陷的} \end{cases} \tag{13-8}$$

其中，i 为第 i 次抽样，j 为 i 次抽样中的第 j 个抽样单元。

则缺陷数 y_i 为

$$y_i = \sum_{j=1}^{n} y_{ij} \tag{13-9}$$

因每次抽样的样本容量为 n，而每次抽样检验的缺陷数为 y_i，通常称它们的比值为每次抽样的缺陷率估值，记为

$$\hat{u}_i = \frac{y_i}{n} \tag{13-10}$$

若 X 的总体容量为 N，总缺陷数为 Y，则 $\hat{u}_i$ 是总体缺陷率 $u=\frac{Y}{N}$ 的估值。

在属性数据总体 X 中，当总体容量 N 足够大时，对该数据总体抽样 m 次，若每次抽样的样本容量为 n，记每个抽样单元的缺陷值为 y_{ij}，一般认为缺陷数 y_{ij} 服从泊淞分布。另外 $E(y_{ij}) = \bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$，其中 y_{ij} 为每个抽样单元的缺陷值，估计量 $\hat{u}_i$ 为

$$\hat{u}_i = \bar{y}_i = \hat{\lambda} = \frac{1}{n}\sum_{j=1}^{n} y_{ij} \tag{13-11}$$

其中，n 为每次抽取的样本容量。

如果第 i 次从总体中抽取 n 条属性数据记录，其中的缺陷总数为 y_i，则 y_i 亦服从泊淞分布，并且，$E(y_i) = n\lambda$；$V(y_i) = n\lambda$，则属性数据的缺陷率的数学期望和方差分别为

$$E(\hat{u}_i) = E\left(\frac{y_i}{n}\right) = \lambda = u \tag{13-12}$$

$$V(\hat{u}_i) = V\left(\frac{y_i}{n}\right) = \frac{\lambda}{n} = \frac{u}{n} \tag{13-13}$$

缺陷率的数学期望是属性数据抽样样本中缺陷多少的度量，缺陷率方差则是属性数

据抽样样本数据中缺陷离散程度的度量。由于缺陷数服从泊淞分布，所以缺陷率数学期望与方差可用于对属性数据质量的度量。

可见，属性数据的缺陷数和缺陷率是通过抽样检验的方法，并以已经颁布的测绘数据质量检验标准或已知值为依据来判断而获得的。它们反映了这些属性数据的可靠程度，也就是反映了属性数据的质量。因此，将它的期望与方差作为属性数据质量的度量既是合理的，也是简单实用的。

采用缺陷率来度量数据的精度，这在质量管理学中，被认为是研究大量非定量数据精度的一种简单且切实可行的方法，而它的前提是要选择费用合理，抽样精度高的抽样方法抽取样本，以获得更好的说明总体数据质量的样本检验结果。一般随机抽样是最基本的抽样方法。但是对于大批量的检验总体，有时总体各部分质量的实际分布差别很大，对不同部分采用不同的检验程序则比较符合实际。所以在 GIS 数据质量的抽样检验中，引入分层概念是符合 GIS 数据组织的基本特点的。

在对 GIS 所有不同数据层分层抽样时，可考虑估计每一层的缺陷率，层的缺陷率估值通过适当加权就能得到对整个总体的缺陷率的估计量。

设 k（$k=1, 2, \cdots, h$）层检验总体的缺陷均率为 U_k，则总体缺陷率的加权平均值为

$$U_w = \sum_{k=1}^{h} W_{\text{layer}(k)} U_k = W_{\text{layer}(1)} U_1 + W_{\text{layer}(2)} U_2 + \cdots + W_{\text{layer}(h)} U_h \quad (13\text{-}14)$$

这里 $W_{\text{layer}(k)}$（$k=1, 2, \cdots, h$）是每一层的权，且有 $\sum_{k=1}^{h} W_{\text{layer}(k)} = 1$，同理可得样本缺陷率估值的加权平均值为：

$$\hat{u}_w = \sum_{k=1}^{h} W_{\text{layer}(k)} \hat{u}_k \quad (13\text{-}15)$$

其中，$\hat{u}_k$ 表示 k（$k=1, 2, \cdots, h$）层样本的缺陷率估值，$\hat{u}_w$ 表示总体样本的缺陷率估值。

检验方差可在各层中分别计算，然后通过加权合并形成样本缺陷率的方差估计量：

$$V(\hat{u}_w) = \sum_{k=1}^{h} W_{\text{layer}(k)}^2 V(\hat{u}_k) \quad (13\text{-}16)$$

通过分层抽样技术，采用缺陷率度量属性数据精度，关键是获得各分层属性数据缺陷的抽样调查数据。而不同的抽样方案对总体抽样获得的结果是不同的，合适的抽样方案应该是在某一费用要求下获得属性数据缺陷的较为准确的调查数据。确定抽样方案实际上是确定各数据层的抽样比，抽样比一方面涉及抽样样本量的大小，另一方面也需要顾及抽样样本对总体的代表性。由于 GIS 属性数据不同数据层的数据特性以及数据规模都有较大的差别，因此，分层抽样比一般抽样能更有效地控制抽样误差。此外，对缺陷的分类可以分别获得对一般缺陷、重缺陷和严重缺陷的缺陷率描述，这三个缺陷率也可以作为属性数据精度的描述内容。

以某一城市土地利用数据作为例子，数据 A 和数据 B 是同一类型的两个空间相邻的数据集。用地类型的划分则是通过 1∶1000 大比例地形图的判读、航片的叠加以及外业调绘等方法获得。由于该区域是农村用地，所以其主要用地类型有四种，分别是渔业

用地、耕地、工矿居住用地和林园用地。

在完成数据采集后，需要对两个数据集属性数据的质量给予度量和评定，由于不同用地类型在整个土地利用中具有不同的土地使用特性，而且属性数据项的数据类型种类很多（表 13-3 给出了土地利用属性数据的检查内容），所以对其质量进行检验和评定需要采用抽样检验的方法，且按四个用地类型进行分层抽样，表 13-4 是其抽样检验结果。

表 13-3　土地利用属性数据检查内容

名　称	层号	用地面积	用地类型	用地权属	用地所属行政区域	用地界址点	产量	承包期限	植被属性	属性记录长度
渔业用地	1	√	√	√	√	√	√	√		7
耕地	2	√	√	√	√	√	√			6
工矿居住用地	3	√	√	√	√	√				5
林、园地	4	√	√	√	√	√	√	√	√	8

表 13-4　样本数据分层抽样检验结果

数　据	层号 k	层权 $W_{layer(k)}$	各层抽样数 n_k	数据记录长度 r_k	一般缺陷数 $y_{kj}^{(1)}$	重缺陷数 $y_{kj}^{(2)}$	严重缺陷数 $y_{kj}^{(3)}$	缺陷数据和 y_{kj}
数据 A	1	0.30	100	7	10	3	1	24
	2	0.20	150	6	15	2	0	21
	3	0.25	200	5	20	1	1	28
	4	0.25	150	8	15	2	1	26
数据 B	1	0.30	200	7	16	5	2	41
	2	0.20	150	6	10	3	0	19
	3	0.25	100	5	5	2	1	16
	4	0.25	200	8	10	3	2	29

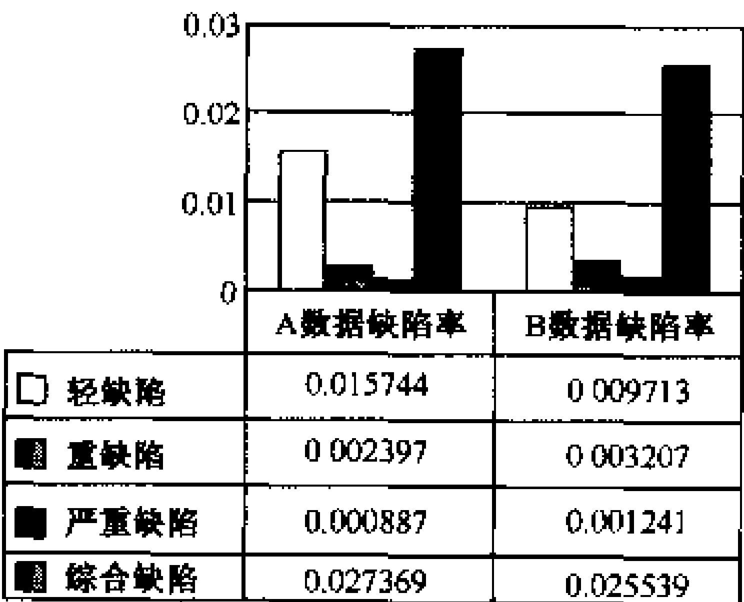


图 13-1　样本数据用不同缺陷类型对属性数据精度的直观描述

可见，这两个数据集根据用地类型分为四个数据层，各层数据的权重在数据采集方案中根据数据的规模以及重要性已经明确给定，同时属性数据的缺陷类型也根据测绘数据质量检验标准划分为一般缺陷、重缺陷和严重缺陷。这样就可以计算两个数据集的属性数据质量的度量值即缺陷率及其估计方差。

在具体度量某一属性数据精度时，需要知道一般缺陷、重缺陷、严重缺陷及综合缺陷的缺陷率，这些缺陷率可以作为属性数据精度的度量。图 13-1 是不同分类的缺陷率度量的直观表达，它可以较容易地度量属性数据的精度。而表 13-5 是这些缺陷率值的估计方差，也即可信度。

表 13-5 样本数据不同缺陷分类的精度度量的方差值

数据集	属性数据精度度量	一般缺陷数 $y_{kj}^{(1)}$	重缺陷数 $y_{kj}^{(2)}$	严重缺陷数 $y_{kj}^{(3)}$	综合缺陷 y_{kj}
数据 A	缺陷率方差 $D(\hat{u})$	1.74E-05	2.93E-06	1.04E-06	3.14E-05
数据 B	缺陷率方差 $D(\hat{u})$	1.09E-05	3.8E 06	1.5E-06	2.98E 05

由上述结果可以得出以下结论：

(1) A 数据的重、严重缺陷率虽然比 B 数据的重、严重缺陷率低，但由于其一般缺陷率远高于 B 数据，所以综合缺陷率反而高了。也就是说，A 数据的总体精度要比 B 数据低，即 B 数据总体质量相对要好一些。可见属性数据质量既可以从不同缺陷类型去评价，也可以综合评价，这样对属性数据质量的度量和控制都更具体。

(2) 缺陷率值的大小本身可作为质量好坏的评定。本例中数据 A 和数据 B 的缺陷率值都非常小，可以认为它们的属性数据绝对质量比较好。但严格地说，应该有一个缺陷率限差，说明属性数据质量合格或不合格。而缺陷率值大小的直接含义需严格依据理论从给定的生产方和用户风险综合考虑，推求缺陷率限差，还需通过大量实际数据分析加以验证。这一理论问题尚需进一步探讨。

(3) 采用缺陷率的方法度量属性数据质量综合考虑了多项数据质量标准，这有别于只研究属性分类精度的误差矩阵和 kappa 系数等度量指标，对于描述属性数据的总体质量是有意义的。

(4) 缺陷率值的取得完全依赖恰当的抽样数据，而抽样数据能否全面地反映整体数据则取决于抽样方案的选取。采用分层抽样方案既考虑了大多数 GIS 数据的存储特征，同时也能降低抽样误差，使得抽样数据能更好的代表整体数据。当然探讨更优化的抽样方案有利于提高缺陷率的可信度。

§13.3 位置误差分析算法

13.3.1 几种误差指标的算法

绝对精度在坐标系中通过控制点给出，精度指标是标准差或平方根，表示观测值与其真值间的差异。需要高质量的地面真值，费用高。相对精度是一个点相对其邻近点的精度，它的确定也需要地面测量，只是区域内测量，并不需要和控制点相连，相对便宜。因为点组成线、线围成面，所以先从点的精度谈起。GIS 中点位的误差，可以用以下几个指标来描述，包括均方根误差、圆概率误差、圆标准误差、误差带的面积、标准差、重心精度、质心精度等（Caspary and Scheuring, 1992）。

1. 点的精度指标

设 GIS 中点位观测值与其真值间的差异为 dx、dy，则它们的均方根误差（Root Mean Square Error）s_x，s_y 可以分别定义为：

$$s_x = \sqrt{\frac{\sum \mathrm{d}x^2}{n}}, \qquad s_y = \sqrt{\frac{\sum \mathrm{d}y^2}{n}} \tag{13-17}$$

由此得到均方根点位位置误差(Root Mean Square Positional Error):

$$\mathrm{RMSP} = \sqrt{s_x^2 + s_y^2} \tag{13-18}$$

圆概率误差(Circular Probable Error):

$$\mathrm{CEP} = 0.59(s_x + s_y) \tag{13-19}$$

圆标准误差(Circular Standard Error):

$$\mathrm{CES} = 0.5(s_x + s_y) \tag{13-20}$$

2. 线的精度指标

Chrisman (1982) 提出的 ε 带，由线段及其平行线围成的区域定义，是线段误差最可能出现的区域，线段的真值在两条 ε 带的中心，ε 带的宽度是点位误差的平方根，在线段的两个端点，为以 ε 为半径的两个半圆。设线段的长度为 l，则 ε 带的面积是

$$A_1 = 2l\varepsilon + \pi\varepsilon^2 \tag{13-21}$$

如果线段的端点坐标误差独立且相等,即标准差

$$\sigma = \sigma_x = \sigma_y \tag{13-22}$$

点 P_i 是线段 P_1P_2 上的任意点,那么根据误差传播律,有 P_i 的点位误差

$$x_i = x_1 + \frac{x_2 - x_1}{l}l_i, \quad y_i = y_1 + \frac{y_2 - y_1}{l}l_i, (0 \leqslant l_i, 0 \leqslant l)$$

$$\sigma_{x_i}^2 = \sigma_{y_i}^2 = \left(1 - 2\frac{l_i}{l} + 2\frac{l_i^2}{l^2}\right)\sigma^2, \tag{13-23}$$

$$\mathrm{RMSP}_i = \sqrt{2}\sigma_{x_i}$$

可见,误差带不是一个长方形,而是一个中部凹陷的线段和真实线段围成的区域,那么 ε 带的面积是

$$A_2 = \pi\varepsilon^2 + 2l\varepsilon\int_0^1 \sqrt{1 - 2x - 2x^2}\,\mathrm{d}x \approx \pi\varepsilon^2 + 1.62l\varepsilon - A_1 - 0.38l\varepsilon \tag{13-24}$$

因为线段中部的点位误差较小,所以此处可用误差带代替 ε 带,又为了便于应用,此处使用 4 条直线逼近误差带,即

$$A_3 = \pi\varepsilon^2 + 4 \times \left[\frac{1}{2} \times \frac{l}{2} \times (1 + 1/\sqrt{2})\varepsilon\right] \approx 1.71l\varepsilon + \pi\varepsilon^2 = A_1 - 0.29l\varepsilon \tag{13-25}$$

可进一步简化为

$$A_4 = 1.6l\varepsilon$$

3. 多边形的精度

多边形(面)由直线段构成,每条线又由其两个端点决定。因此,多边形(面)的精度就和前面讨论的点和线的统计特性密切相关。面的标准差根据误差传播律的特性计算,其面积公式是

$$2A = \sum_{i=1}^{n}(y_{i+1} - y_{i-1})x_i, \quad y_0 = y_n, y_{n+1} = y_1,$$

$$\sigma = \sigma_x = \sigma_y = \mathrm{RMSP}_i / \sqrt{2} \tag{13-26}$$

对于所有的坐标 x, y,多边形(面)的标准差是:

$$\sigma_A = \sigma \sqrt{2A\sin(360°/n)} = \mathrm{RMSP} \sqrt{A\sin(360°/n)} \tag{13-27}$$

$n=3$:三角形,

$$A = \frac{\sqrt{3}}{4}a^2,$$

$$\sigma_A = \mathrm{RMSP}\sqrt{A\sqrt{0.75}} = \mathrm{RMSP}\cdot a \cdot 0.6124$$

$n=4$:四边形,

$$A = a^2,$$

$$\sigma_A = \mathrm{RMSP}\sqrt{a^2} = \mathrm{RMSP}\cdot a$$

$n=8$:多边形,周长 $u=8a$,

$$A = 4.8284a^2,$$

$$\sigma_A = \mathrm{RMSP}\cdot a \cdot 1.8478$$

重心精度(Accuracy of the center of gravity)

$$x_c = \frac{1}{n}\sum x, \sigma_{x_c} = \frac{1}{\sqrt{n}}\sigma,$$

$$y_c = \frac{1}{n}\sum y, \sigma_{y_c} = \frac{1}{\sqrt{n}}\sigma, \tag{13-28}$$

$$\mathrm{RMSP}_c = \sqrt{\frac{2}{n}}\sigma$$

质心精度(Accuracy of the center of mass)

$$c = \sum_{i=1}^{n}\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2} = \sum_{i=1}^{n}a, i = 1,2,\cdots,n; n+1 \text{ 取值为 } 1$$

$$\sigma_c^2 = \frac{\sigma^2}{a^2}\cdot\sum_{i=1}^{n}[(\Delta x_i - \Delta x_{i+1})^2 + (\Delta y_i - \Delta y_{i+1})^2]$$

$$\Delta x_i = x_{i+1} - x_i, \Delta y_i = y_{i+1} - y_i \tag{13-29}$$

13.3.2 置信区域方法

一般地讲,GIS 中所描述的量测点与其在现实世界中的真实位置并不完全一致。这是因为量测值含有误差。其真实位置通常是位于某一量测位置附近的某一区域。本节给出了 GIS 特征位置不确定性的统计模型。该模型描述了该区域的特征;包括其几何位置及形状。对于二维线段或线不确定性的研究主要有以下三个方面:①线段上某一点的误差;②线段分布的概率描述;③置信域。线段上任意一点之误差可以用误差传播定律的方法导出。Shi (1994) 对此作了研究,得到的结果是线段中点之误差为两端点误差的 0.707 倍。为研究线段或线的概率分布,Dutton (1992) 用 MonteCarlo 模拟法模拟了其分布。而文献 Shi (1994) 则从概率论与数量统计的角度在理论上对分布问题予以证明。但是以上讨论仅限于二维 GIS 中几何特征的位置不确定性。本节则对 GIS 中不确定性

问题进行更一般地扩展：从一维、二维到 N 维空间。

1. GIS 中的几何特征

GIS 中的一个几何特征可以是一维的，如一个点的高程。在目前的 GIS 应用中二维空间的特征则最为普遍。在空间问题为三维的情况下，时空 GIS 则涉及四维 GIS 的问题。为了使讨论更一般化，本节介绍了 N 维 GIS 空间的概念。由于一、二维 GIS 几何特征比较直观，首先讨论了这两维 GIS 中几何特征的不确定性问题。进一步地，不确定问题又被一般化，即讨论了 N 维空间 GIS 之不确定性。在每一维空间内，GIS 的几何特征可以被划分为点、线段和线。在描述几何特征的位置不确定性时，多边形可以被看成是一种特殊的线——封闭的线。因此，本节讨论的范围定为一维、二维和 N 维空间内点、线段、和线特征的位置不确定性。一维、二维 GIS 中的几何特征如图 13-2 所示

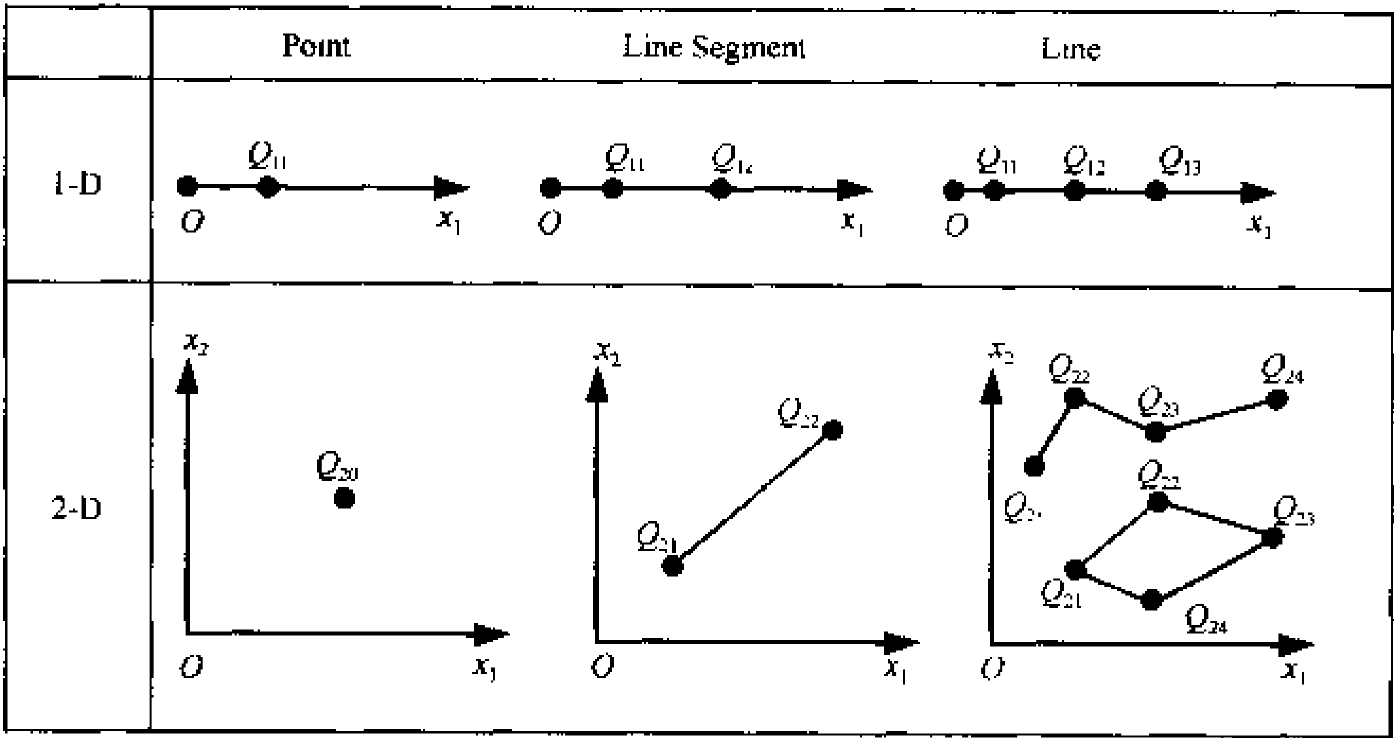


图 13-2 GIS 中的一维与二维几何特征

接着将讨论一维到 N 维 GIS 几何特征的位置不确定性。一般认为，在一维 GIS 中的点误差均符合正态分布，并且假设点之间的误差是相互独立的。

2. 一维特征的位置不确定性

1）一维点

在一维空间内，一个点可以定义为一个一维随机矢量：$Q_{10}=[X_{10}]$，并且服从一维正态分布(N_1)：

$$Q_{10}=[X_{10}]\sim N_1[\mu_{10},\sigma_{11}]$$

这里 μ_{10} 和 σ_{11} 是确定该点统计特性的参数。量测点 Q_{10} 是对其真值 $\Phi_{10}=[\mu_{10}]$ 的一个估计量。

一维点的置信区间 J_{10} 是一个包含了量测值 Q_{10} 真值的区间，真值以大于某一预先定义的置信水平(γ)的概率被包含在内，即：

$$P(\Phi_{10} \in J_{10}) > \gamma$$

其中 J_{10}是满足以下条件的一组点集$[x_1]$

$$X_{10} - a_{11} \leqslant x_1 \leqslant X_{10} + a_{11}$$

这里

$$a_{11} = k^{1/2}(\sigma_{11})^{1/2}$$

参数 k 与预先定义的置信水平 γ 有关。γ 可以从 χ^2 分布表中查得，$k = K_{1;\gamma}^2$。例如，当 $\gamma = 0.97$时，$k = 4.789$。其结果如图 13-3 所示。

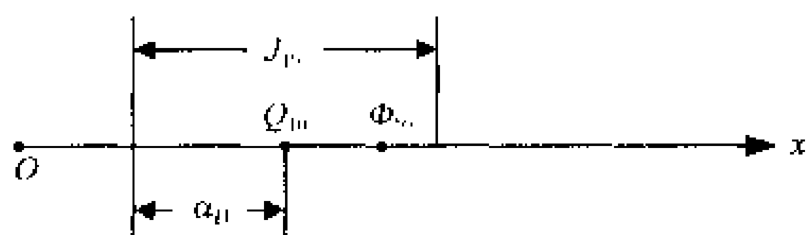


图 13-3 一维点的置信区间

2）一维线段

在一维空间中，两点确定一线段，并且由两个一维随机矢量 $Q_{11} = [X_{11}]$和 $Q_{12} = [X_{12}]$组成，同时，假设它们服从一维正态分布：

$$Q_{11} = [X_{11}] \sim N_1[\mu_{11}, \sigma_{11}]$$

$$Q_{12} = [X_{12}] \sim N_1[\mu_{12}, \sigma_{11}]$$

这里 μ_{11}，μ_{12}和 σ_{11}是决定这两点统计特性的参数。量测点 Q_{11}和 Q_{12}是两端点真值 μ_{11}和 μ_{12}的量测值。

一条线段可以看成是由一端点到另一端点连续移动的任意点组成的。线段 $\Phi_{11}\Phi_{12}$间的任意点可以表示为：

$$\Phi_{1r} = (1-r)\Phi_{11} + r\Phi_{12} = [(1-r)\mu_{11} + r\mu_{12}], \quad r \in [0,1] \qquad (13\text{-}30)$$

(1) 线段上任意点的分布

根据上式，线段上任意点为

$$Q_{1r} = (1-r)Q_{11} + rQ_{12} = [(1-r)X_{11} + rX_{12}] = [X_{1r}]$$

由于线段上的任意一点 Q_{1r}是两端点之随机矢量 Q_{11}和 Q_{12}的线性函数，因此 Q_{1r}亦服从一维正态分布：

$$Q_{1r} \sim N_1\{[(1-r)\mu_{11} + r\mu_{12}]\}, \{(1-r)^2 + r^2\sigma_{11}\}$$

(2) 一维线段的置信区间

由于一维线段两端点存在量测误差，线段的真实位置一般与其量测值($Q_{11}Q_{12}$)不重合，但被包含在该线段附近的一个区间内。我们把这个区间定义为置信区间(J_1)，并且导出了该置信区间的数学表达式(Shi, 1994)。该推导主要利用了线段上任意点的统计分布及 χ^2 分布，其结果如下。

一维线段 Q_1Q_2 的置信区间是这样一个区间，所有点 Φ_{1r}(其中 $\gamma \in [0,1]$)以大于预先定义的一个置信水平 γ 的概率被包含于J_{1r}内，即：

$$P(\Phi_{1r} \in J_{1r}, \text{对于所有的 } r \in [0,1]) > \gamma$$

J_1 是所有 $J_{1r}(r\in[0,1])$的并集。一个 J_{1r}是满足以下条件的 x_1 的集合

$$X_{1r}-C_{11}\leqslant x_1\leqslant X_{1r}+C_{11}$$

这里

$$C_{11}=k^{1/2}[((1-r)^2+r^2)\sigma_{11}]^{1/2}$$

参数 k 与选定的置信水平有关，并且可以从 x^2 分布表中查得，$k=K^2_{2;\gamma}$，例如，若 $\gamma=0.97$，则 $k=7.714$。

在置信水平表达的风险下，我们认为，线段的真实位置($\Phi_{11}\Phi_{12}$)在围绕量测线段所构成的置信区间内。该置信区间是一条一维的线段，如图 13-4 所示。

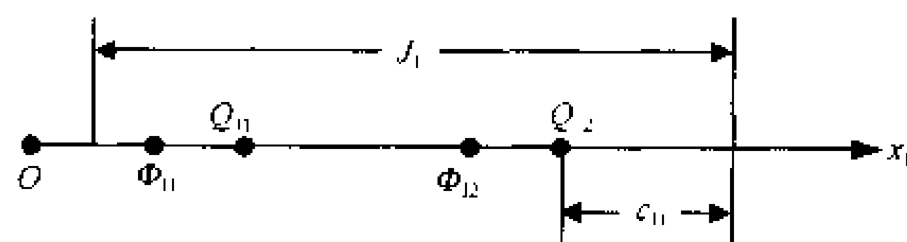


图 13-4　一维线段的置信区间

(3) 一维线的置信区间

在一维空间内，设一条线(Q_1Q_m)是由一系列的线段构成：Q_0Q_1，Q_1Q_2，…，$Q_{m-1}Q_m$。根据线段置信区间的概念，可以分别获得各线段的置信区间，$I_1,I_2,I_3,\cdots,I_m$。由此可以获得线 Q_1Q_m 的置信区间，即为所构成线段的置信区间的并集：

$$I_{1m}=\bigcup_{j=1}(I^m_{(j-1)(j)})\qquad(j=1,2,\cdots,m)$$

该结果用图 13-5 表示。设所有线段的置信区间均是在同一置信水平下获得的，例如 97%，在这种情况下，线 Q_1Q_m 置信区间的置信水平亦为任一线段的置信水平，即 97%。

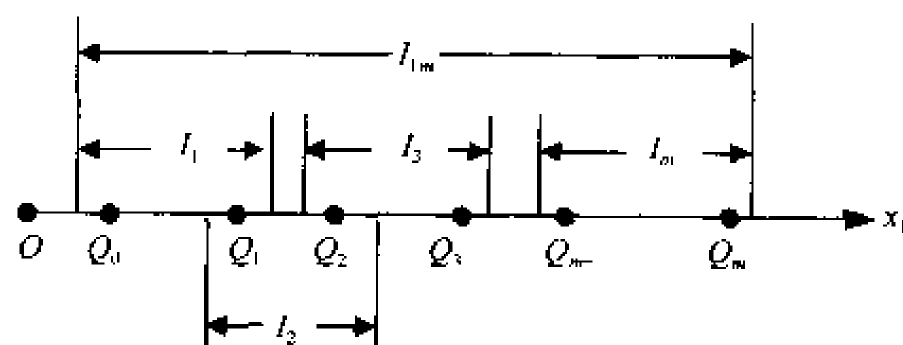


图 13-5　一条一维线的置信区间

3. 二维几何特征的位置不确定性

1) 二维点

在一个二维空间内，一个点被定义为一个二维量测的随机矢量 $Q_{20}=[X_{10},X_{20}]^T$ 并且服从以下二维正态分布(N_2)：

$$Q_{20}=\begin{vmatrix}X_{10}\\X_{20}\end{vmatrix}\sim N_2\left[\begin{vmatrix}\mu_{10}\\\mu_{20}\end{vmatrix},\begin{vmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{vmatrix}\right]$$

这里 μ_{10}，μ_{20}，σ_{11}，σ_{22}和 $\sigma_{12}=\sigma_{21}$是决定量测量 Q_{20}统计特性的参数。Q_{20}是其真值

$\Phi_{20}=[\mu_{10},\mu_{20}]^T$的一次估计量。

如果一个二维点 Q_{20}的方差、协方差阵已知，我们可以导出其统计分量的 χ^2 分布以及点 Q_{20}的置信域 J_{20}。

二维量测点 Q_{20}的置信域被定义为这样一个区域，其真值以大于某一预先定义的置信水平(γ)的概率落在该置信域 J_{20}内，即：

$$P(\Phi_{20}\in J_{20})>\gamma$$

置信域 J_{20}是由以下一个点集$[x_1,x_2]^T$ 构成的，其中 x_1 满足：$X_{10}-a_{21}\leqslant x_1\leqslant X_{10}+a_{21}$并且 x_2 满足：$X_{20}-a_{22}\leqslant x_2\leqslant X_{20}+a_{22}$。这里：$a_{21}=k^{1/2}(\sigma_{11})^{1/2}$，并且：$a_{22}=k^{1/2}(\sigma_{22})^{1/2}$

参数 k 与选定的置信水平 γ 有关，并且可以从一个 χ^2 分布表中查得，$k=K^2_{1;(1+\gamma)/2}$，例如，如果 $\gamma=0.97$，则$(1+r)/2=0.985$，$k=6.024$。其结果如图 13-6 所示。一个二维点的置信域是一个矩形域，其大小取决于量测点的方差 σ_{11}、σ_{22}以及预先定义的置信水平 γ。

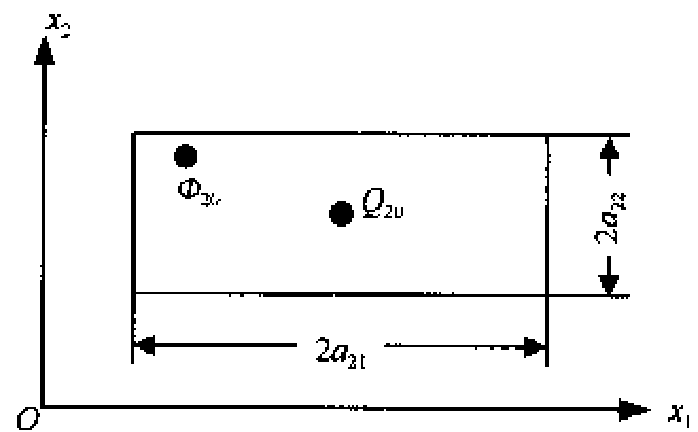


图 13-6 二维点的置信域

2）二维线段

在一个二维空间内，我们定义一条线段是由其两端点的两个二维随机矢量 $Q_{21}=[X_{11},X_{21}]^T$ 和 $Q_{21}=[X_{12},X_{22}]^T$ 构成的，并且满足以下两个二维正态分布(N_2)：

$$Q_{21}=\begin{vmatrix}X_{11}\\X_{21}\end{vmatrix}\sim N_2\left[\begin{vmatrix}\mu_{11}\\\mu_{21}\end{vmatrix},\begin{vmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{vmatrix}\right]$$

$$Q_{22}=\begin{vmatrix}X_{12}\\X_{22}\end{vmatrix}\sim N_2\left[\begin{vmatrix}\mu_{12}\\\mu_{22}\end{vmatrix},\begin{vmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{vmatrix}\right]$$

这里 $\mu_{11},\mu_{21},\mu_{12},\mu_{22},\sigma_{11},\sigma_{22},\sigma_{12}=\sigma_{21}$是确定两个随机矢量的统计参数。量测点 Q_{21}和 Q_{22}是其真值 $\Phi_{21}=[\mu_{11},\mu_{21}]^T$ 和 $\Phi_{21}=[\mu_{12},\mu_{22}]^T$ 的估计值。

一条二维线段可以表示为连接两端点的一个连续的点集。线段真值 $\Phi_{21}\Phi_{22}$上的任意一点 Φ_{2r}可以表示为：

$$\Phi_{2r}=(1-r)\Phi_{21}+r\Phi_{22}=\begin{vmatrix}(1-r)\mu_{11}+r\mu_{12}\\(1-r)\mu_{21}+r\mu_{22}\end{vmatrix}\quad r\in[0,1]\qquad(13\text{-}31)$$

(1) 二维线段上任意一点的分布

根据式(13-31)，量测线段上任意点为：

$$Q_{2r}=(1-r)Q_{21}+rQ_{22}=\begin{vmatrix}(1-r)X_{11}+rX_{12}\\(1-r)X_{12}+rX_{22}\end{vmatrix}=\begin{vmatrix}X_{1r}\\X_{2r}\end{vmatrix}$$

该任意点服从以下二维正态分布：

$$Q_{2r}\sim N_2\left[\begin{vmatrix}(1-r)\mu_{11}+r\mu_{12}\\(1-r)\mu_{21}+r\mu_{22}\end{vmatrix},|(1-r)^2+r^2|\begin{vmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{vmatrix}\right]$$

(2) 二维线段置信域

该线段的真值可以表达为 Φ_{2r}的一个点集，$r\in[0,1]$。

二维线段的置信域 J_2 是这样一个区域，所有 $\Phi_{2r}(r\in[0,1])$被以大于某一预先定义的置信水平的概率包含于该置信域内，即：

$$P(\Phi_{2r}\in J_{2r},\text{对于所有 } r\in[0,1])>\gamma \tag{13-32}$$

J_2 是所有 $J_2(r\in[0,1])$的并集。一个 J_{2r}是一个点集$(x_2,x_2)^T$ 构成的，其中 x_1 满足：

$$X_{1r}-C_{21}\leqslant x_1\leqslant X_{1r}+C_{21} \tag{13-33}$$

并且 x_2 满足

$$X_{2r}-C_{22}\leqslant x_2\leqslant X_{2r}+C_{22} \tag{13-34}$$

这里

$$C_{21}=k^{1/2}\|(1-r)^2+r^2|\sigma_{11}|^{1/2} \tag{13-35}$$

且

$$C_{22}=k^{1/2}\|(1-r)^2+r^2|\sigma_{22}|^{1/2} \tag{13-36}$$

参数 k 与置信水平 γ 有关，$k=K^2_{2(1+r)/2}$。例如，$\gamma=0.97$，$(1+\gamma)/2=0.985$，$k=8.517$。

(3) 二维线段置信域的形状

对式 13-35 和 13-36 求偏导。在$\partial C_{21}/\partial r=0$ 和$\partial C_{22}/\partial r=0$ 的条件下，$r=0.5$ 时取得其最小值。C_{21}和 C_{22}的另外两个极值是当 $r=0$ 或 $r=1$ 时。即置信域在其两端点取得极大值而在中间取得极小值，如图 13-7 所示。该形状与 Epsilon 带的形状有所不同。Epsilon 带是在各处大小均匀的一个平行带。因此，Epsilon 带是对本文所提置信区间的真实描述的一种近似。

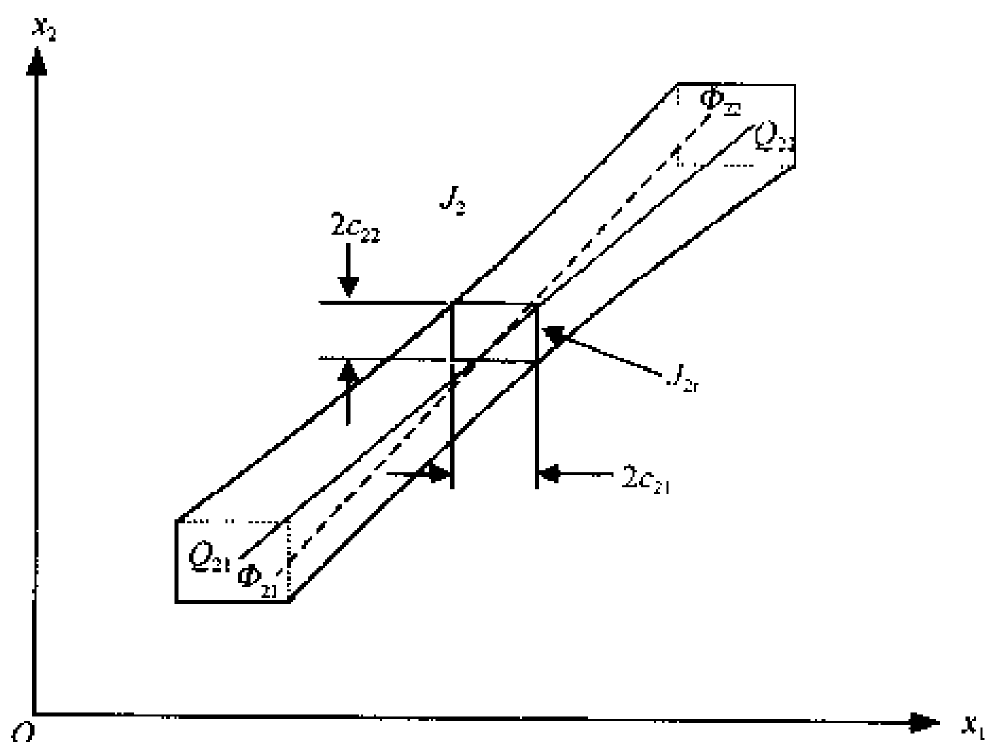


图 13-7 二维线段的置信域

根据置信区间的解析表达式(13-32)～(13-36)，可知置信区间的大小取决于二维线段两端点的误差及预先定义的置信水平。两端点的误差及置信水平越大，置信域的面积也越大。置信域的形状是中间小、两端大的一个条带状区域。

3）二维线的置信域

设一条二维线（Q_1Q_m）是由一系列二维线段（$Q_0Q_1, Q_1Q_2, \cdots, Q_{m-1}Q_m$）所构成的，该线段的置信域分别为 $R_1, R_2, \cdots, R_m$。其定义为量测线周围的一个区域以大于预先给定的置信水平的概率包含了该线的真实位置。线 Q_1Q_m 的置信域是所有构成该线段的置信域的并集，即：

$$R_{1m} = \bigcup_{j=1} \{R^m_{(j-1)(j)}\} \qquad (j = 1, 2, \cdots, m)$$

图 13-8 所示为线 Q_1Q_m 的置信区域。设所有线段的置线域是在同一置信水平（例如 97%）下获得的，则该二维线置信域的置信水平亦为该值，即 97%。

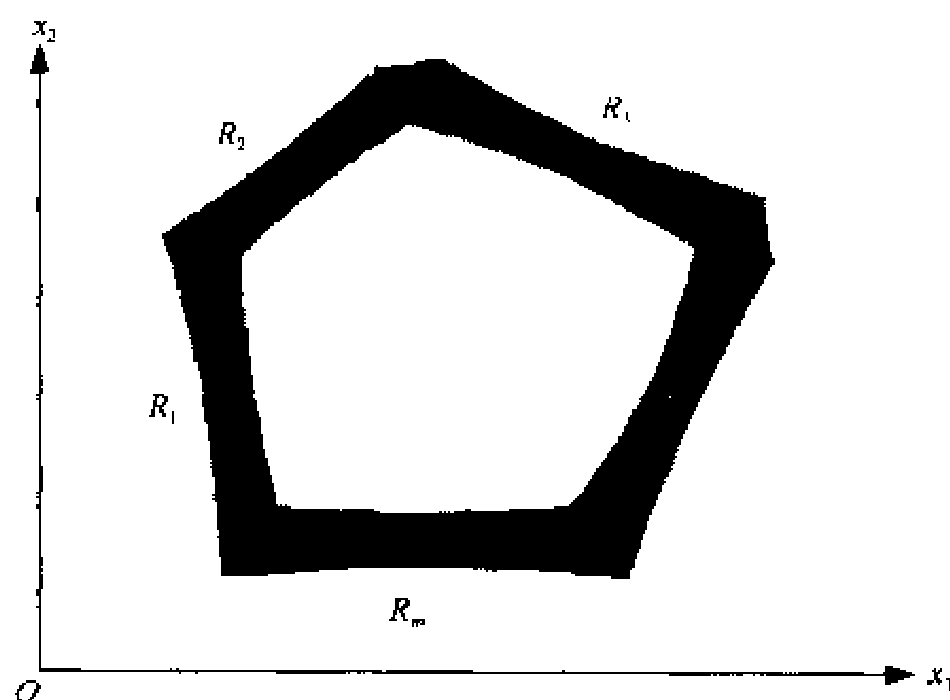


图 13-8　一条二维线的置信域

4. N 维空间几何特征的不确定性

1）N 维点

在一个 N 维空间内，一个点被定义为一个 N 维随机矢量 $Q_{N0} = [X_{10}, X_{20}, \cdots, X_{N0}]^T$，它服从 N 维正态分布（N_N）：

$$Q_{N0} = \begin{bmatrix} X_{10} \\ X_{20} \\ \vdots \\ X_{N0} \end{bmatrix} \sim N_N \left[\begin{vmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{N0} \end{vmatrix}, \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{vmatrix} \right]$$

这里 $\mu_{i0}, \sigma_{ii}, \sigma_{ij} = \sigma_{ji} (i, j = 1, 2, \cdots, N)$ 是决定随机矢量统计特性的参数，量测量 Q_{N0} 是其真值 $\Phi_{N0} = [\mu_{10}, \mu_{20}, \cdots, \mu_{N0}]^T$ 的一个估计量。

为描述 N 维空间内点的不确定性，我们导出一个子空间——置信空间。

N 维随机点 Q_{N0} 的置信空间（J_{N0}）是这样一个空间，量测矢量（Q_{N0}）的真值（Φ_{N0}）以大于某一预先定义的置信水平（γ）的概率被包含在该空间内，即

$$P(\Phi_{N0} \in [J_{N0}]) > \gamma$$

J_{N0}是一个点集$[x_1, x_2, \cdots, x_N]^T$。其中 x_i 满足

$$X_{i0} - a_{Ni} \leqslant x_i \leqslant X_{i0} + a_{Ni} \quad (i = 1,2,\cdots,N)$$

这里 $a_{Ni} = k^{1/2}(\sigma_{ii})^{1/2}(i = 1,2,\cdots,N)$

其中,$k = K^2_{1;(N-1+\gamma)/N}$可以从$\chi^2$分布表中表得。在置信水平所在的风险下,可以认为量测 N 维点的真值被包含在该置信体内。

2) N 维线段

定义一条 N 维线段的两个 N 维随机矢量 $Q_{N1} = [X_{11}, X_{21}, \cdots, X_{W1}]^T$ 和 $Q_{N2} = [X_{12}, X_{22}, \cdots, X_{W2}]^T$,且满足 N 维正态分布(N_N):

$$Q_{N1} = \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{N1} \end{bmatrix} \sim N_N \left[\begin{vmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{N1} \end{vmatrix}, \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{vmatrix} \right]$$

$$Q_{N2} = \begin{bmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{N2} \end{bmatrix} \sim N_N \left[\begin{vmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{N2} \end{vmatrix}, \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{vmatrix} \right]$$

这里 $\mu_{i1}, \mu_{i2}, \sigma_{ii}, \sigma_{ij} = \sigma_{ji}(i,j = 1,2,\cdots,N)$是决定两端 N 维矢量统计特性的参数。一条 N 维线段是由其上的任意点的集合组成的。$\Phi_{N1}\Phi_{N2}$上任意一点 Φ_{Nr}可以表示为:

$$\Phi_{Nr} = (1-r)\Phi_{N1} + r\Phi_{N2} = \begin{vmatrix} (1-r)\mu_{11} + r\mu_{12} \\ (1-r)\mu_{21} + r\mu_{22} \\ \vdots \\ (1-r)\mu_{N1} + r\mu_{N2} \end{vmatrix}, r \in [0,1] \qquad (13\text{-}37)$$

(1) N 维线段上任意点的分布

根据式(13-37),相应的量测线段上的任意点可表示为:

$$Q_{Nr} = (1-r)Q_{N1} + rQ_{N2} = \begin{vmatrix} (1-r)X_{11} + rX_{12} \\ (1-r)X_{21} + rX_{22} \\ \vdots \\ (1-r)X_{N1} + rX_{N2} \end{vmatrix} = \begin{vmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{Nr} \end{vmatrix}$$

由于 Q_{Nr}是Q_{N1}、Q_{N2},并且 Q_{N1}和 Q_{N2}相互独立且服从 N 维正态分布。因此,Q_{wr}服从以下正态分布:

$$Q_{Nr} \sim N_N \left[\begin{vmatrix} (1-r)\mu_{11} + r\mu_{12} \\ (1-r)\mu_{21} + r\mu_{22} \\ \vdots \\ (1-r)\mu_{N1} + r\mu_{N2} \end{vmatrix}, |(1-r)^2 + r^2| \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{vmatrix} \right]$$

(2) N 维线段的置信空间

N 维线段的置信空间(J_N)是这样一个空间,该线段上所有点 Φ_{Nr}($r \in [0,1]$以大于某一预先确定的置信水平(γ)的概率被包含于该空间内,即:

$$P(\Phi_{Nr} \in J_{Nr}, \text{对于所有 } r \in [0,1]) > \gamma$$

J_N是J_{Nr}集合的并集($r \in [0,1]$)。一个 J_{Nr}是一个点的集合$(x_1, x_2, \cdots, x_N)^T$。其中 $x_i = (i = 1, 2, \cdots, N)$满足:

$$X_{ir} - C_{Ni} \leqslant x_i \leqslant X_{ir} + C_{Ni}$$

这里

$$C_{Ni} = k^{1/2} \| (1-r)^2 + r^2 | \sigma_{ii} |^{1/2} \qquad (i = 1, 2, \cdots, N)$$

其中 $k = K^2_{2;(N-1+\gamma)/N}$ 可以从 χ^2 分布表中查得。置信空间的大小与两端点的误差及选定的置信水平有关。

3) N 维线的置信空间

N 维线的置信空间定义为这样一个子空间,量测线的真值以大于某一置信水平的概率落在该区间内。设 N 维线 Q_1Q_m 是以一组线段 Q_0Q_1, Q_1Q_2, …, $Q_{m-1}Q_m$ 组成的。每一构成线段的置信空间分别是 S_1, S_2, …, S_m。根据线的置信空间的定义,我们可以得到线 Q_1Q_m 置信空间的表达式:

$$S_{1m} = \bigcup_{j=1}^{m} \{ S_{(j-1)(j)} \} \quad (j = 1, 2, \cdots, m)$$

若所有线段均取相同的置信水平(97%),则该 N 维线置信空间的置信水平亦为97%。

广义的 N 维位置为不确定模型中的“位置”,具有几何特征并且遵从几何规则。例如 N 为 3 时即为三维模型。N 维模型的实际意义在于解决多维 GIS 中的不确定性问题。例如 N 为 4 时,即可直接用于时—空 GIS 中不确定性的描述。

13.3.3 误差概率分布

研究线元的不确定性模型的主要问题之一是认识线元误差的分布状况,给出描述方法,并恰当地建立 GIS 中线元的误差带缓冲区。在这方面的研究中,Perkal (1956) 等建立的‘ε－带’模型取 ε 为常数值,认为线元的真位置将以一定的概率落在此,ε－带中;Chrisman (1982) 和 Caspary (2000) 引用 Perkal (1956) 提出的‘ε－带’的概念开始描述矢量 GIS 中的线元位置不确定性;Dutton (1992) 等用模拟法描述了线元位置不确定性的分布;Shi (1994, 1998) 提出线元的线段上任意点以在 x 和 y 方向上同时所在的置信区间证明了线段的置信域及其概率分布;而对‘ε－带’模型的扩展是以线元上各点在其垂直方向的中误差(标准差)或‘最大方向中误差’表示 ε－带宽(刘大杰 1999);有的学者将线元上的点看做随机过程,用随机过程的概率密度来描述 ε－带。此外,线元不确定性的可靠带(刘文宝 1995)和信息熵误差带(范爱民 2001)也相继用来探讨 GIS 线元不确定性模型的有关问题。

对一维正态分布的随机误差 $\Delta \sim N(0, \sigma)$ 落在区间 $(-k\sigma, k\sigma)$ 内的概率的分析,有效地描述了正态误差的分布状况;对平面点位落在其误差椭圆内的概率的分析,也有效地描述了平面点位的分布状况。因此,在给出 GIS 线元不确定性的描述方法,

也希望能借助于概率分析来描述和说明线元误差的分布状况，这也是人们十分关注的问题。Shi（1998，2000）探讨了线元的误差指标和置信区域问题，应用 Cauchy-Schwarz 不等式，得到当置信水平为 95%时，线元两个自由度置信区域的宽度小于 $2\sqrt{2\times 5.99}\sigma_0$。

这里，将平面线元看做是由无限个相互接近的点构成的随机过程，认为平面线元落在某一区间内的概率，可以通过这些点在线元垂直方向的误差曲面构成的体积得到。还采用数值积分方法计算了平面线元落在不同区间内的概率。并以此为依据，进一步讨论了平面线元的误差指标和误差带缓冲区的选取问题。考虑到平面线元误差的概率是在平面点位误差的概率的基础上计算，因此，还对平面点位误差落在几个不同区间内的概率作了讨论。

1. 平面点点位误差的概率

平面控制点的点位精度通常用误差椭圆来描述，设任意点 A 的坐标为(x,y)，它们的方差和协方差为 $\sigma_x^2,\sigma_y^2,\sigma_{xy}$；$\rho_{xy}$是$x$ 与 y 的相关系数，其联合概率密度函数为：

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$$

$$\exp\left\{-\left(\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\frac{1}{2(1-\rho_{xy}^2)}\right\} \quad (13\text{-}38)$$

可得到误差椭圆元素(长、短半轴 E、F 及长半轴的方位角θ)为：

$$E=\frac{1}{2}\sqrt{(\sigma_x^2+\sigma_y^2)+|(\sigma_x^2-\sigma_y^2)^2-4\sigma_{xy}|}$$

$$F=\frac{1}{2}\sqrt{(\sigma_x^2+\sigma_y^2)-|(\sigma_x^2-\sigma_y^2)^2-4\sigma_{xy}|} \quad (13\text{-}39)$$

$$\tan(2\theta)=\frac{2\sigma_{xy}}{\sigma_x^2-\sigma_y^2}$$

且可证明平面点落在误差椭圆范围 Ω 内(如图 13-9)的概率就是二维误差曲面在 Ω 内的体积：

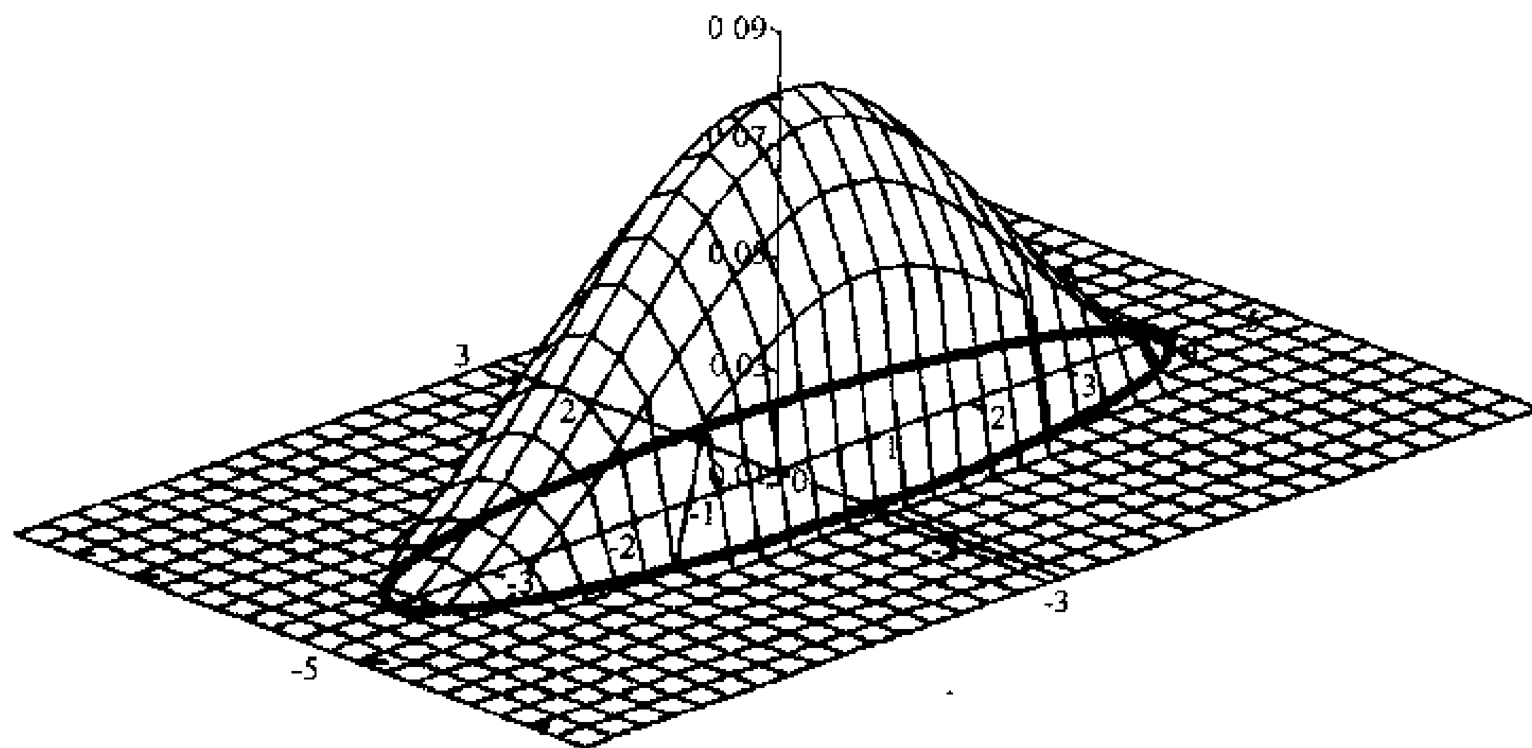


图 13-9 二维误差曲面落在误差椭圆内

$$P_r(x, y \subset \Omega) = \iint_{\Omega} f(x, y)\mathrm{d}x\mathrm{d}y = 4\int_0^r\int_0^{\frac{\pi}{2}} f(r, \theta)\mathrm{d}r\mathrm{d}\theta = 0.393 \tag{13-40}$$

而二倍和三倍误差椭圆范围 Ω 内的概率分别为 0.893 和 0.989。

为了进一步描述点位误差的分布状况，下面讨论平面点落在不同区间内的概率。可以证明，以误差椭圆的长、短半轴为坐标轴，A 点在该坐标系中的坐标为 (u, v)，则 u, v 的相关系数等于 0，其联合概率密度为

$$\begin{aligned} f(u, v) &= f_u(u)f_v(v) = \frac{1}{\sqrt{2\pi}E}\exp\left(-\frac{u^2}{2E^2}\right)\frac{1}{\sqrt{2\pi}F}\exp\left(-\frac{v^2}{2F^2}\right) \\ &= \frac{1}{2\pi EF}\exp\left(-\frac{u^2}{2E^2}-\frac{v^2}{2F^2}\right) \end{aligned} \tag{13-41}$$

A 点落在 $(-E<u<E; -F<v<F)$ 范围内的概率，就是该二维误差曲面在误差椭圆外切矩形区域内的体积（如图 13-10）

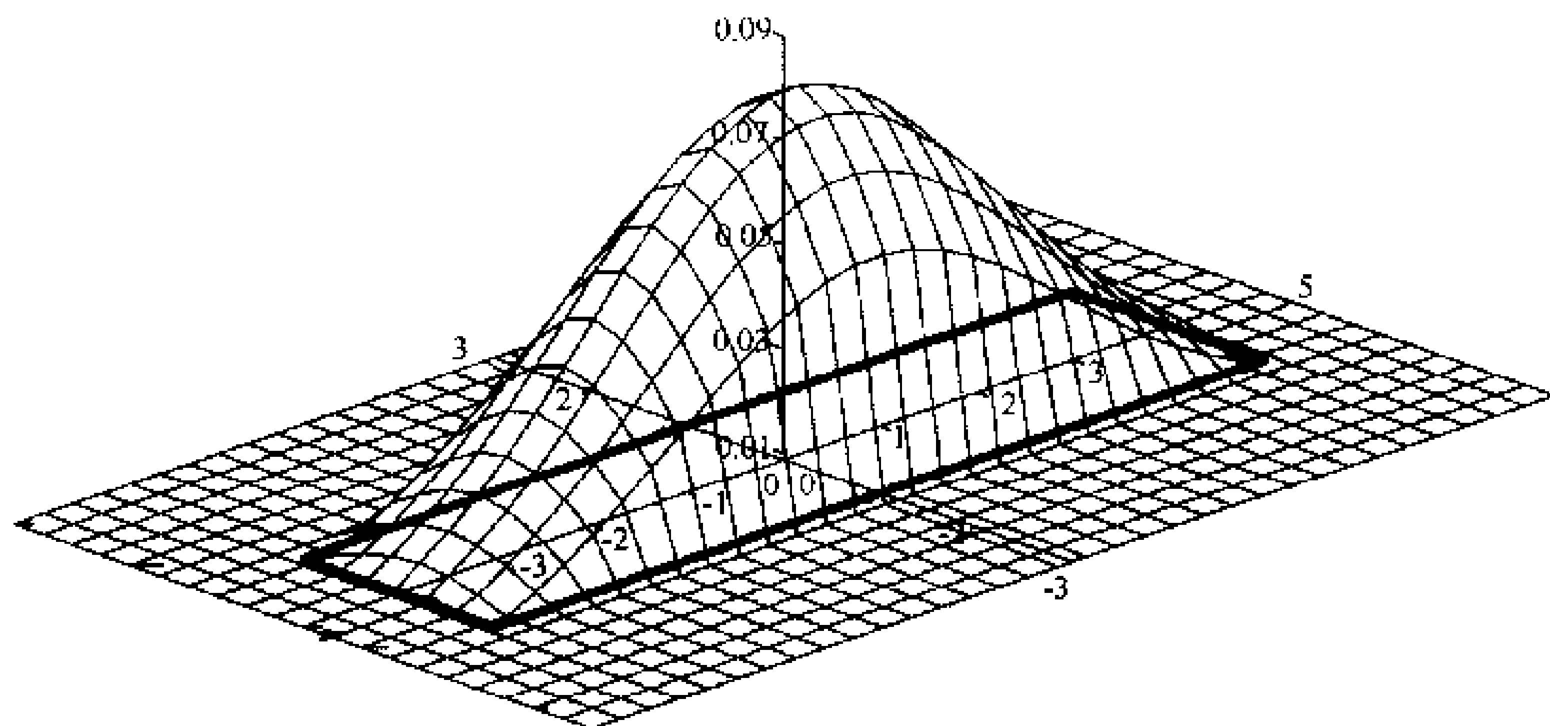


图 13-10　二维误差曲面落在误差椭圆外切矩形内

$$\begin{aligned} P_r(-E<u<E; -F<v<F) &= 4\int_0^E\int_0^F \frac{1}{2\pi EF}\exp\left(-\frac{u^2}{2E^2}-\frac{v^2}{2F^2}\right)\mathrm{d}u\mathrm{d}v \\ &= 2\int_0^E \frac{1}{\sqrt{2\pi}E}\exp\left(-\frac{u^2}{2E^2}\right)\mathrm{d}u \times 2\int_0^F \frac{1}{\sqrt{2\pi}F}\exp\left(-\frac{v^2}{2F^2}\right)\mathrm{d}v \\ &= 0.683\times 0.683 = 0.466 \end{aligned} \tag{13-42}$$

同样可得，落在 $(-2E<u<2E; -2F<v<2F)$ 和 $(-3E<u<3E; -3F<v<3F)$ 范围内的概率分别为

$$P_r(-2E<u<2E; -2F<v<2F) = (0.954)^2 = 0.9111$$

$$P_r(-3E<u<3E; -3F<v<3F) = (0.997)^2 = 0.9946$$

A 点在原坐标系中的坐标为 (x, y)，联合概率密度为 (13-38) 式。点位落在 $\Omega_x(-\sigma_x<x<\sigma_x; -\sigma_y<y<\sigma_y)$ 范围内的概率为（如图 13-11）

$$P_r(-\sigma_x<x<\sigma_x, -\sigma_y<y<\sigma_y) = \iint_{\Omega_x} f(x, y)\mathrm{d}x\mathrm{d}y \tag{13-43}$$

当 $\sigma_x=\sigma_y=1$，相关系数 ρ_{xy} 分别为 0.3、0.5 和 0.7 时，计算得到 P_r 分别为 0.5096、0.5642 和 0.6657。

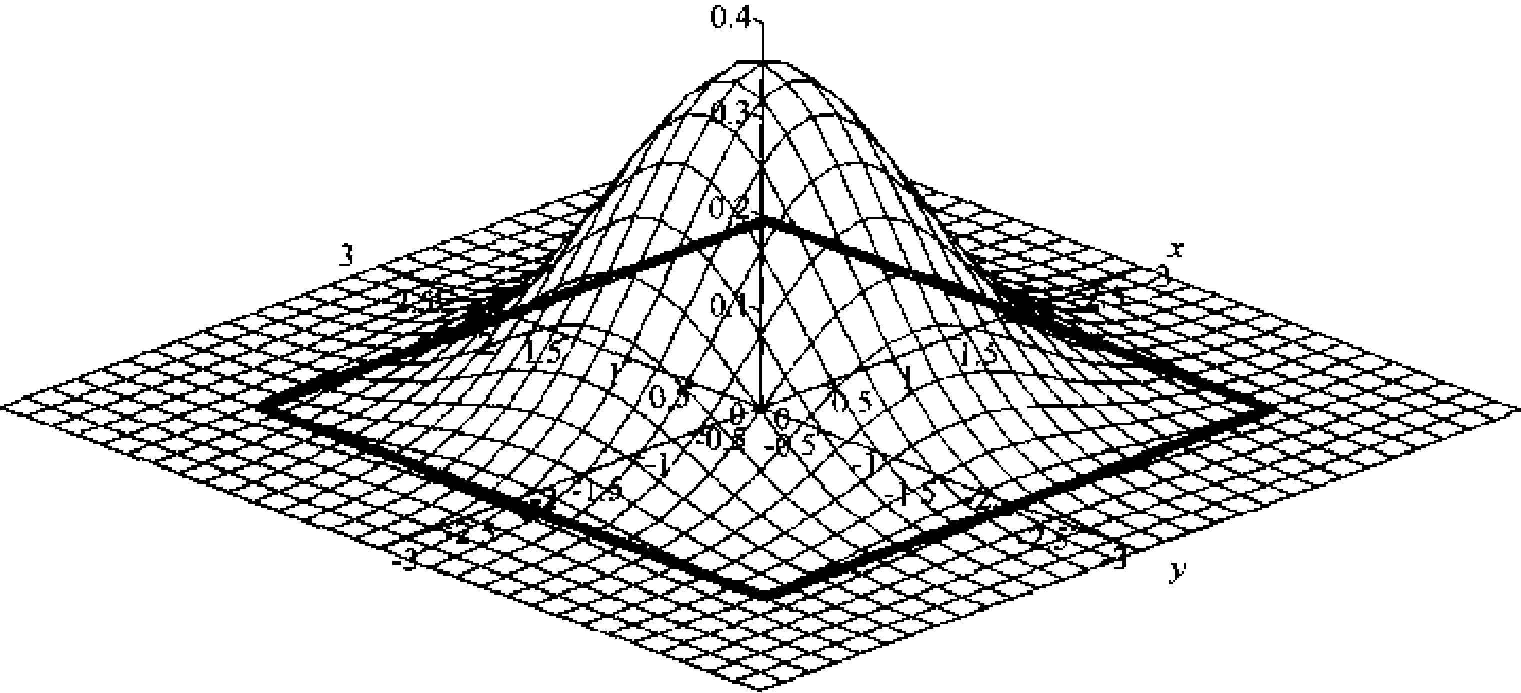


图 13-11　二维误差曲面落在 Ω_x 矩形内

从上述概率值可见，平面点而落在（$-\sigma_x<x<\sigma_x$；$-\sigma_y<y<\sigma_y$）范围内的概率随相关系数的不同而不同。而落在误差椭圆范围 Ω 内的概率为 0.393，落在（$-E<u<E$；$-F<v<F$）的矩形区域内的概率是 0.466，都是固定值。因此，通常用落在误差椭圆范围 Ω 内的概率 0.393 或落在矩形区域（$-E<u<E$；$-F<v<F$）内的概率 0.466 来描述平面点位误差的分布状况。

2. 平面线元的误差带及其概率

设线元二端点 A 与 B 的坐标为 $A(x_1,y_1)$，$B(x_2,y_2)$，它们的协方差阵为：

$$D=\begin{bmatrix}\sigma_{x1}^2 & \sigma_{x1y1} & \sigma_{x1x2} & \sigma_{x1y2}\\ \sigma_{x1y1} & \sigma_{y1}^2 & \sigma_{y1x2} & \sigma_{y1y2}\\ \sigma_{x1x2} & \sigma_{y1x2} & \sigma_{x2}^2 & \sigma_{x2y2}\\ \sigma_{x1y2} & \sigma_{y1y2} & \sigma_{x2y2} & \sigma_{y2}^2\end{bmatrix}$$

线元上的任意点 $A_r(x_r,y_r)$，设 AB 的距离为 l，AA_r 的距离为 l_r，$r=l_r/l$（如图 13-12），则 $A_r(x_r,y_r)$ 的方差和协方差为：

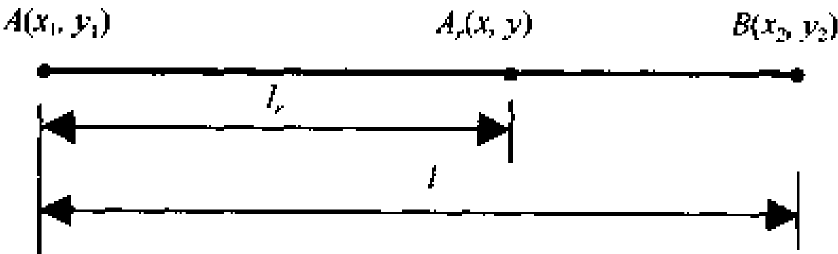


图 13-12　线元上点的表达

$$\sigma_{xr}^2=(1-r)^2\sigma_{x1}^2+2r(1-r)\sigma_{x1x2}+r^2\sigma_{x2}^2$$

$$\begin{aligned}\sigma_{xryr} &= (1-r)^2\sigma_{x1y1} + r(1-r)(\sigma_{y1x2} + \sigma_{x1y2}) + r^2\sigma_{x2y2} \\ \sigma_{yr}^2 &= (1-r)^2\sigma_{y1}^2 + 2r(1-r)\sigma_{y1y2} + r^2\sigma_{y2}^2\end{aligned} \tag{13-44}$$

$A_r(x_r, y_r)$的概率密度为：

$$f(x_r, y_r) = \frac{1}{2\pi\sigma_{xr}\sigma_{yr}\sqrt{1-\rho_r^2}}\exp\left\{-\frac{1}{2(1-\rho_r^2)}\left[\frac{x_r^2}{2\sigma_{xr}^2} + \frac{(y_r - l_r)^2}{2\sigma_{yr}^2} - \frac{2\rho_r x_r(y_r - l_r)}{\sigma_{xr}\sigma_{yr}}\right]\right\} \tag{13-45}$$

若 $\sigma_{x1x2} = \sigma_{x1y2} = \sigma_{y1y2} = \sigma_{x1y1} = \sigma_{y1x2} = \sigma_{x2y2} = 0$

$\sigma_{xr}^2 = (1-r)^2\sigma_{x1}^2 + r^2\sigma_{x2}^2$

$\sigma_{xryr}^2 = 0$

$\sigma_{yr}^2 = (1-r)^2\sigma_{y1}^2 + r^2\sigma_{y2}^2$

则式(13-45)的概率密度变为

$$f(x_r, y_r) = \frac{1}{2\pi\sigma_{xr}\sigma_{yr}}\exp\left\{-\frac{x_r^2}{2\sigma_{xr}^2} - \frac{(y_r - l_r)^2}{2\sigma_{yr}^2}\right\} \tag{13-46}$$

在 AB 上取点 $A_1, A_2, \cdots, A_n$，它们的联合概率密度可写为

$$f(X) = f(x_1, y_1, \cdots, x_n, y_n) = \frac{1}{(2\pi)^{1/2}|D_x|}\exp\left\{-\frac{1}{2}(X - \mu_x)^T D_x^-(X - \mu_x)\right\} \tag{13-47}$$

其中 X 是由 n 个点的坐标构成的向量，D_x，μ_x 是相应的方差和期望。

根据多维随机变量分布函数的性质，可以得到线元上各点同时落在 $\Omega(-\sigma_{xr} < x_r < \sigma_{xr}; -\sigma_{yr} < y_r < \sigma_{yr})$内的概率是

$$P_r = \int_\Omega f(X)\mathrm{d}X = \int_\Omega \frac{1}{(2\pi)^{1/2}|D_x|}\exp\left\{-\frac{1}{2}(X - \mu_x)^T D_x^-(X - \mu_x)\right\}\mathrm{d}X \tag{13-48}$$

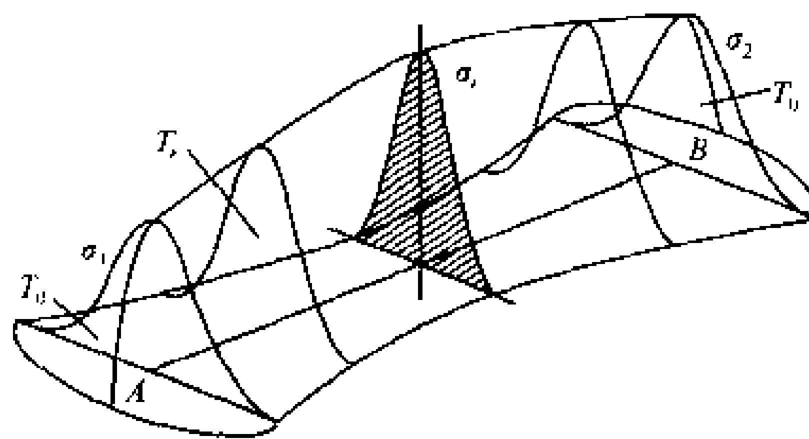


图 13-13　平面线元误差带的概率示意图

当这些点无限接近时，线元上各点同时落在 Ω 内的概率可以认为就是图 13-13 所示图形体积的一个比值，有

$$P_r = \frac{T_r + T_0}{T_l + 1} \tag{13-49}$$

其中，T_r 是线元上除端点外各点在垂直于线元方向的误差曲面所构成的在区间 $\Omega_r(-\varepsilon < x_r < \varepsilon; 0 < y_r < l)$内的体积；$T_l$ 是相应的在区间 $\Omega_0(-\infty < x_r < \infty; 0 < y_r < l)$内的体

积，T_0 是线元的两个端点的点位误差曲线构成的体积。若取：

$$\varepsilon = \sigma_{xr} = \sqrt{(1-r)^2\sigma_{x1}^2 + r^2\sigma_{x2}^2} \tag{13-50}$$

由于 $y_r = l_r$，则：

$$\begin{aligned} T_r &= 2\int_0^{\sigma_r}\int_0^l f(x_r)\mathrm{d}x_r\mathrm{d}y_r = 2\int_0^{\sigma_r}\int_0^l f(x_r)\mathrm{d}x_r l\mathrm{d}r \\ &= 2\int_0^l\int_0^l \frac{l}{\sqrt{2\pi}}\exp\left\{-\frac{x_r^2}{2}\right\}\mathrm{d}x_r\mathrm{d}r = 0.6826l \end{aligned} \tag{13-51}$$

也可得

$$T_l = 2\int_0^{\infty}\int_0^l \frac{l}{\sqrt{2\pi}\sigma_{xr}}\exp\left\{-\frac{x_r^2}{2\sigma_{xr}^2}\right\}\mathrm{d}x_r\mathrm{d}r = l \tag{13-52}$$

对于 T_0，如果取其误差椭圆范围的体积，则 $T_{01} = 0.393$，如果取（$-E < u < E$；$-F < v < F$）矩形区间的体积，则 $T_{02} = 0.466$，这样可得到线元在误差带区域内的概率如表 13-6。

表 13-6 线元误差带的概率

P_r \ l	$l=5\sigma_x$	$l=10\sigma_x$	$l=100\sigma_x$	$l=500\sigma_x$	端点
P_{r1}	0.6345	0.6565	0.6799	0.6822	($T_{01}=0.393$)
P_{r2}	0.6433	0.6613	0.6805	0.6823	($T_{02}=0.466$)

可见，对于不同的 ε 区间，有不同的概率。从式(13-48)～(13-52)和表 13-6 的结果还可知，线元上的点落在（$-\sigma_{xr} < x_r < \sigma_{xr}$；$-\sigma_{yr} < y_r < \sigma_{yr}$）内的概率 P_r 与在线元二端点所取的误差范围（误差椭圆或矩形区间）有关，因而也与线元的长度 l 有关，当 $l = 5\sigma_x$，端点取误差椭圆范围或矩形区间，分别得 $P_{r1} = 0.6345$，$P_{r2} = 0.6433$，而当 $l = 500\sigma_x$，则得 $P_{r1} = 0.6822$，$P_{r2} = 0.6823$。即当 $l \gg \sigma_x$，P_r 趋近于 68.3%。

参 考 文 献

范爱民，郭达志，2001．误差熵的不确定性模型．测绘学报，30(1)：48～53

李德仁，袁修孝，2002．误差处理与可靠性理论．武汉大学出版社，235～295

刘 春，2000．GIS属性数据的精度度量及质量控制的抽样原理与方法[博士学位论文]．上海：同济大学

刘大杰，史文中等，1999．GIS空间数据的精度分析与质量控制．上海：上海科技文献出版社

刘文宝，1995．GIS数据的不确定性理论．武汉测绘科技大学出版社

刘文宝，邓敏，夏宗国，2000．矢量GIS中属性数据的不确定性分析．测绘学报，29(1)：76～81

史文中，1998．空间误差处理的理论和方法．北京：科学出版社

Burrough P. A., 1999. GIS and Geostatistics: Essential partners for spatial analysis. In Proc. of the International Symposium on Spatial Data Quality'99, 10～20

Caspary W. and Scheuring R., 1992. Error-bands as measures of geometrical accuracy. In EGIS'92 Conference Proceedings of the third European Conference on Geographical Information Systems, 226～233

Caspary W. and Scheuring R., 2000. A Positional Accuracy in Spatial Database Environment and Urban Systems, 1993 (17): 103～110

Chrisman N R., 1982. A theory of Cartographic Error and its Measurement in Digital Database. Auto-Carto, 5, 159～168

Dutton G., 1992. Handling Positional Uncertainty in Spatial Database. In Proceeding of 5th International Symposium on Spatial Data Handling, 460～469

Mikhail E. M. and Ackermann F., 1976. Observations And Least Squares. (New York: IEP-A Dun-Donnelley Publisher), 497

Perkal J., 1956. On Epsilon Length. Bulletin de l' Academic Polonaise Des Sciences, (4): 399～403

Shi W. Z., 1994. Model Positional and Thematic Uncertaintes in Integration of Remote Sensing and GIS. ITC Publicaiton, Enschede, 147

Shi W. Z. and Liu W. B., 2000. A stochastic process-based model for the positional error of line segments in GIS. Int. J of Geographical Information Science, 14:51～66

Zhang J., 1996. A Surface-based Approach to Handling Uncertainties in an Urban orientated Spatial Database. Ph.D Thesis (Edinburgh: The University of Edinburgh)

第十四章　多边形自动生成与裁剪算法

多边形图是GIS的基本图形类型，常用来描述面状分布的地理要素。这种图形的重要特点是每个图斑都是一个连通区域，有明确的界线。区域和界线是多边形图的两个基本要素，界线由边界点和结点组成，区域、界线、结点之间的拓扑关系是GIS的基本理论问题之一。如何根据源数据正确、自动、快速地生成多边形集合及其间的拓扑关系，是GIS算法研究的一项重要任务，直接关系到GIS的功能和效率，是面域空间分析（如多边形叠置分析、缓冲区分析、网络分析）的基础，对空间数据处理也有重要意义。以下从二维角度讨论多边形生成的数学基础、弧-弧拓扑生成算法、多边形自动生成算法和多边形图裁剪算法等基本问题。

§14.1　多边形性质及有关处理

14.1.1　多边形图的基本性质

平面上一条不自相交的有向封闭曲线所形成的图形为多边形，该曲线即为多边形的边界。按左手法则，若边界的前进方向左侧为多边形区域，则该方向为多边形边界的正向。一条封闭曲线有两个方向，分别代表两个不同多边形的边界。平面上由一条不自相交的封闭曲线构成的图形称为最简多边形图，最简多边形图包含两个多边形，一个为封闭曲线的内部区域，另一个为封闭曲线与图形外部边框（边界）之间的区域，或称含有平面上的无穷远点。

从最简单多边形图上引有限条互不相交（可以相切）且不与已有的任何曲线相交（可以相切）的自封闭曲线，所围成的图形称为多边形图，所引的曲线称为弧段，定义了方向的弧段称为有向弧段。显然，多边形图中的每条弧段由且仅由两个多边形共有。

若多边形的边界 L 由曲线 $y=f(x)$ 表示，则有下式表示的积分：

$$A=-\oint_L f(x)\mathrm{d}x \tag{14-1}$$

当 $A>0$ 时，A 为该多边形的面积；当 $A<0$ 时，表示多边形的区域含有平面上的无穷远点。

若多边形的边界 L 的坐标串顺序（按 L 的方向）为：$(x_1,\ y_1)$、$(x_2,\ y_2)$、…、$(x_n,\ y_n)$，且首尾两点坐标相同，则式（14-1）可写为：

$$A=-\frac{1}{2}\sum_{i=1}^{n-1}(y_{i+1}+y_i)(x_{i+1}-x_i) \tag{14-2}$$

由于一幅多边形图的多边形拓扑信息中，每条弧段出现且仅出现两次，两次使用弧段的不同方向，所以，一幅完整的多边形图上，有向弧段的总数是弧段总数的两倍，拓扑信息中记录的弧段次数也是弧段总数的两倍。

14.1.2 异常处理

悬线、桥和无邻接弧段是多边形生成及多边形拓扑关系形成之前必须处理的三类典型异常情形。

数据采集尤其是地图数字化过程中，由于采集误差往往导致曲线在交点处过伸或未及，形成所谓的悬线。悬线的判定一般有两种方法（陈春等，1996）：

方法一：若某一结点仅对应一条弧，则该弧为悬线；

方法二：曲线求交之后，某一曲线的第一交点之前的弧段和最后交点之后的弧段为悬线。

在多边形生成和拓扑信息产生之前，必须先删除悬线。

有时，由于数据采集过程中遗漏了某些弧段，导致某些多边形不封闭，这些不封闭的多边形的某些边可能成为桥。如图 14-1 所示，由于数字化时漏输了弧段 EF，则弧段 AB、CD 都是桥。为了保证多边形生成和拓扑关系建立的正确，必须对桥进行处理，即进行弧段的补入，使桥非桥化。

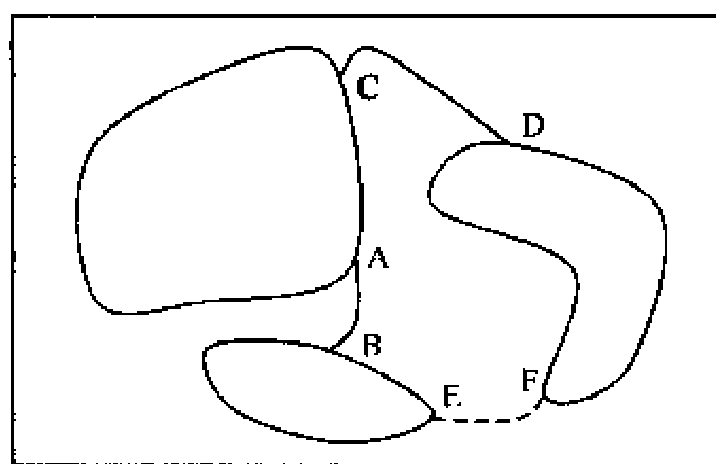


图 14-1 含桥的多边形图示例

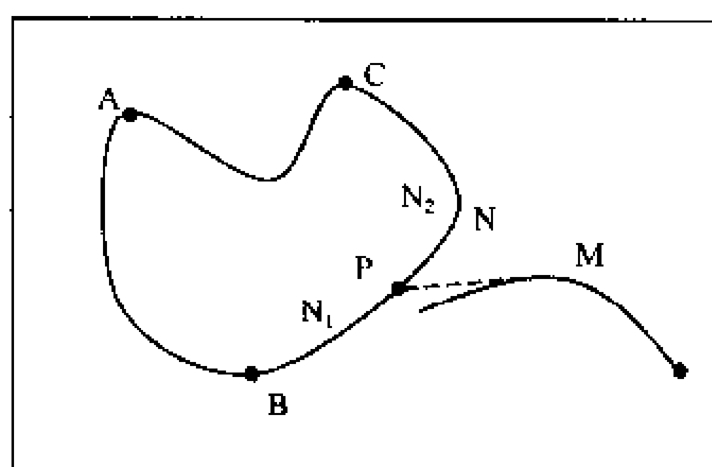


图 14-2 无邻接弧段的处理

有时，由于采样误差和匹配失误，出现部分弧段的首末点与其他弧段无邻接关系。为此，需要确定一个阈值 D 作为端点坐标匹配限差，来实现对弧段的断开处理。设图中全部弧段的最大、最小 X、Y 坐标分别为 X_{max}、X_{min}和 Y_{max}、Y_{min}，设 L 为(X_{max} X_{min}) 和（$Y_{max}-Y_{min}$）中的较小者，A 为最短弧段的长度。根据实验，闫浩文等(2000) 等建议，D 可以取为 $L/1000$ 和 A 的较小者。如图 14-2 所示，若某一弧段 M 的首（末）端点与其他弧段的首（末）端点及本弧段的末（首）端点的距离均大于 D，则搜索其他弧段的坐标，找出与该点距离最近的点 P 及其所在的弧段 N，然后把弧段 N 以 P 为界分成两段，并把 P 点作为弧段 M 的新端点。从而使得弧段 M 与弧段 N_1、N_2 相邻。

14.1.3 岛的判断

首先引入“连通域”概念：若某一多边形内的任意两点之间都有一条位于本多边形内的路径，则该多边形称为连通域。若连通域中的任一闭合曲线可以不经过连通域的边

界而逐渐收缩为一点，则该连通域称为单连通域，否则称为复连通域，如图 14-3 所示。

在含有复连通域的多边形图中，复连通域具有内外两条边界弧段，按左手法则，其中内边界弧段的正向对应复连通域的内部，负向对应的区域称为岛。图 14-3（b）中的多边形 P_2 即为岛。岛的自动识别是多边形拓扑信息生成过程中的重点和难点。

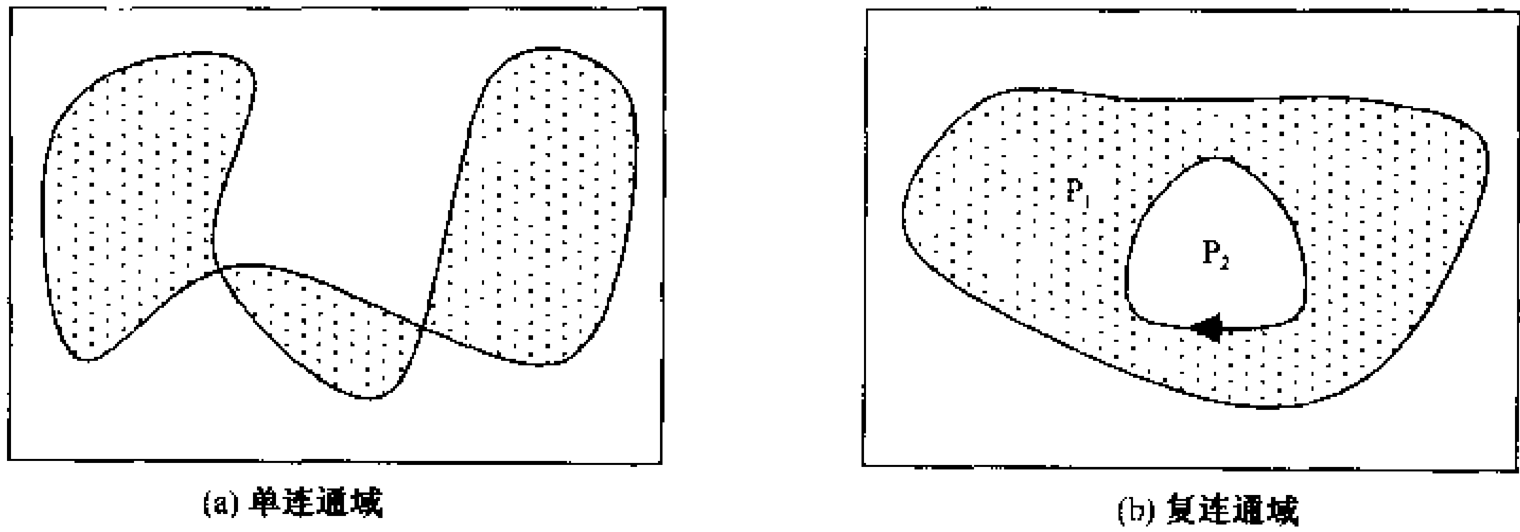


图 14-3　连通域的两种类型

对于既定的弧段方向和边界点排序，当且仅当 $A<0$ 时，多边形为岛。岛有两种类型：其一位于某一外多边形中，其二位于含无穷远点的多边形中，如图 14-4 所示。

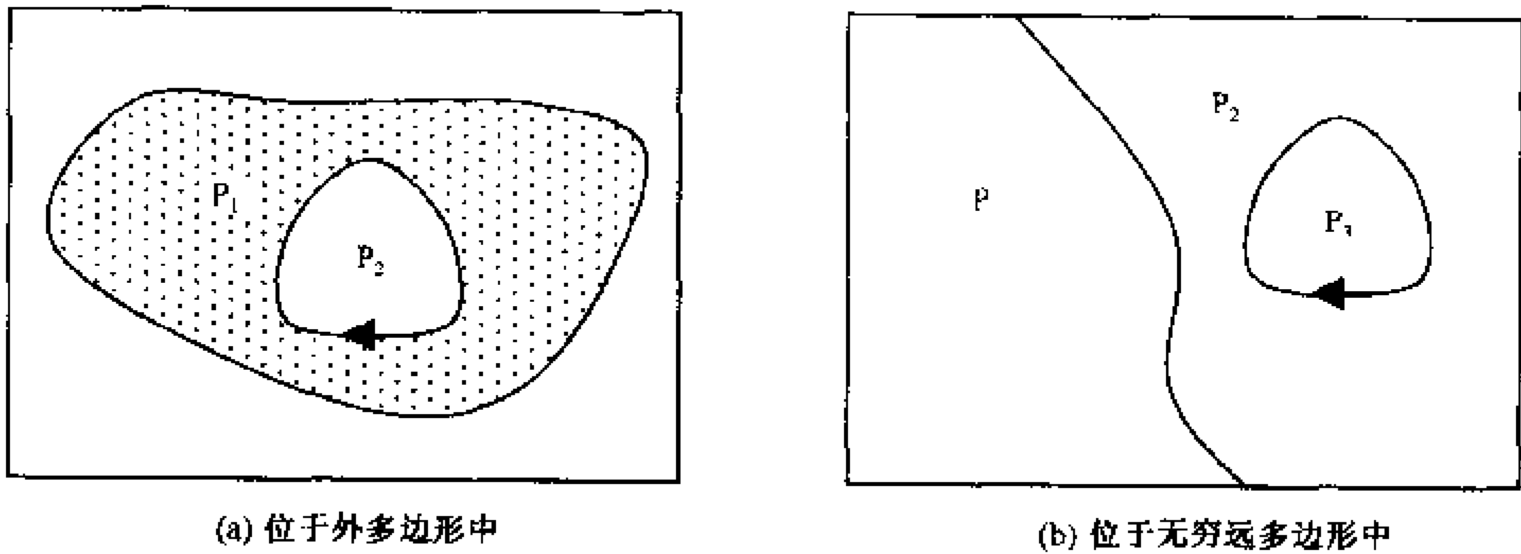


图 14-4　岛的两种基本类型

岛与其外多边形的对应关系是多边形拓扑的一项重要内容。确定这一关系的基本思想为：从岛区多边形的边界点集中任取一点，利用点与多边形的包含关系的判断算法，若该点位于某一非岛区多边形内，则该非岛区多边形即为本岛的外多边形。

14.1.4　内点计算

地理实体多边形的标识是利用多边形的内点来实现的。通过内点将多边形的集合信息和属性信息进行关联。基于手扶跟踪数字化仪采集地图数据的多边形拓扑关系建立，其算法都需要事先手工输入多边形的内点，即在多边形形成之后，用鼠标在多边形内取一点，然后利用点在多边形内的判断算法判断该点是否位于多边形之内；若不在，则继

续进行。显然，这一方法需要大量的人工干预，但与繁杂的数字化过程相比并不突出。而随着地图扫描数字化的普遍应用，多边形内点的输入问题就凸现出来了，一方面是手工输入效率低下，另一方面是不利于寻找和消除矢量化过程产生的无意义的微小多边形（齐华，1997）。

针对这一问题，人们研究提出了一些改进方法。这些改进方法用特定的算法来自动生成多边形的内点，基本做法是以某种方法产生一个点，再用垂线法判断该点是否位于多边形之内（钟振祥，1991）。以质心算法为例，对于凸多边形，质心肯定在多边形内；而对于凹多边形，则不一定，需要进行判断和人工移位。显然，这种算法具有试探性，在不利条件下需要经过多次试探才能生成一个内点，影响效率。齐华（1997）提出基于多边形的外切矩形和多边形内包含所有内边界的最小矩形，利用垂线求交排序原理来直接生成多边形的内点。蔡绍华（1998）对这一基于最小外切矩形（minimum bounding rectangle，MBR）的算法进行了软件实现。此处对这一算法作简单介绍。

基本原理：

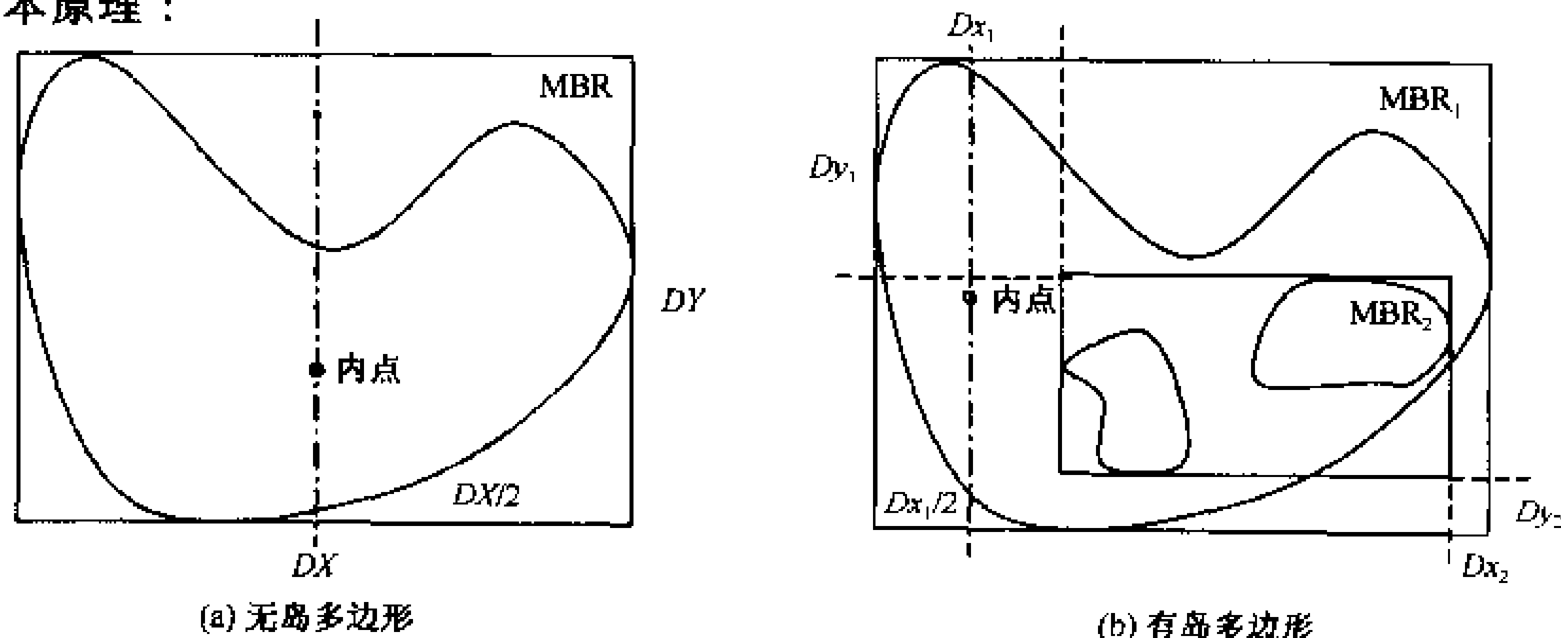


图 14-5　岛的两种基本类型

1）对于无岛多边形（如图 14-5a 所示）

首先，产生多边形的外切矩形 MBR；

其次，比较 MBR 在 X、Y 方向的度量 DX、DY，若 $DX > DY$，则在 $DX/2$ 处作垂直于 X 轴的直线，求取该直线与多边形的交点的 Y 坐标序列，并对 Y 序列由大到小排队（Y_1，Y_2，…，Y_n），然后依次奇偶配对（如 Y_1 与 Y_2，Y_3 与 Y_4……）。

然后，计算奇偶配对后每对的间距，找出其中间距最大者，设为 Y_i 与 Y_{i+1}，则多边形的内点坐标为：

$$\{(X_{\min} + X_{\max})/2, (Y_i + Y_{i+1})/2\} \tag{14-3}$$

2）对于有岛多边形（如图 14-5b 所示）

首先，产生多边形的外切矩形 MBR_1 和多边形内包含所有岛屿的最小矩形 MBR_2；

其次，比较 MBR_1、MBR_2 之左上角、右下角在 X、Y 方向的差异度量 Dx_1、Dy_1 和 Dx_2、Dy_2，求出其中的最大值；如图 14-5b 所示，最大值为 Dx_1，则 $Dx_1/2$ 处作垂

直于 x 轴的直线，求取该直线与多边形的交点的 Y 坐标序列，并对 Y 序列由大到小排队和奇偶配对。

然后，找出奇偶配对后间距最大者为 Y_i 与 Y_{i+1}，则多边形的内点坐标为：

$$\{(\mathrm{MBR}_1(X_{\min}) + \mathrm{MBR}_2(X_{\min}))/2, (Y_i + Y_{i+1})/2\} \tag{14-4}$$

显然，该算法计算的多边形内点具有惟一性解，对于凹、凸多边形均适用，有利于多边形的拓扑维护、拓扑查询和提高效率。

§14.2 弧-弧拓扑生成算法

建立结点上弧-弧拓扑关系的算法直接影响多边形自动生成的时间效率，是多边形自动生成算法优化的关键。建立同一结点上弧-弧拓扑关系实质上是确定弧段之间的相互位置关系，即弧段之间的相对位置关系。传统的算法选择以角度 α_i 为参数，通过 $\tan^{-1}$ （x） 直接求 α_i 或利用泰勒级数的展开式近似计算角度 α_i，进而对 α_i 比较排序来确定弧段之间的相对位置关系。以 $\tan^{-1}$ （x） 直接求 α_i 的算法存在操作繁琐的缺陷，而利用泰勒级数的展开式近似计算角度 α_i 虽然计算效率得到提高，但泰勒级数展开比较复杂，其计算精度也不是很高。近年，齐华等（1996）放弃以角度为参数的传统思想，提出以 Q_i（x，y）函数为参数的新算法，使得计算的时间效率得到较大提高。高云琮等（2002）则认为 Q_i（x，y）算法仍然需要进行比值运算和弧段所在象限的判断，进而从计算几何思想出发提出了一种基于矢量外积的高效算法。

14.2.1 基于方位角的算法

测量学中，坐标方位角是指从坐标北方向起，顺时针旋转到某一射线间的角度。现借用坐标方位角的概念，规定从平面直角坐标系的 X 轴的正半轴方向起，逆时针旋转到某一射线的角度称为该射线的计算方位角，其取值范围为 0°～360°。

设射线 OA 的端点坐标为 $O(x_o, y_o)$、$A(x_A, y_A)$，在 X、Y 轴方向上两端点坐标差为 $D_x = x_A - x_o$，$D_y = y_A - y_o$，则其计算方位角 α_{OA} 的可能取值为：

$$\alpha_{OA} = \begin{cases} 90^\circ \cdots D_x = 0, D_y > 0 \\ 270^\circ \cdots D_x = 0, D_y < 0 \\ \arctan(D_y/D_x) \cdots D_x > 0, D_y \geqslant 0 \\ 360^\circ - \arctan(D_y/D_x) \cdots D_x > 0,\ D_y < 0 \\ 180^\circ - \arctan(D_y/D_x) \cdots D_x < 0 \\ 180^\circ + \arctan(D_y/D_x) \cdots D_x < 0,\ D_y \leqslant 0 \\ \text{不存在} \cdots\cdots D_x = 0, D_y = 0 \end{cases} \tag{14-5}$$

如图 14-6 所示，若从 O 点出发的射线除 OA 之外，还有 OB、OC……，则按上述算法分别计算其方位角 α_{OA}、α_{OB}、α_{OC}……，这些方位角由小到大进行排序，即可以确定从结点 O 出发的各弧段的相对位置关系。

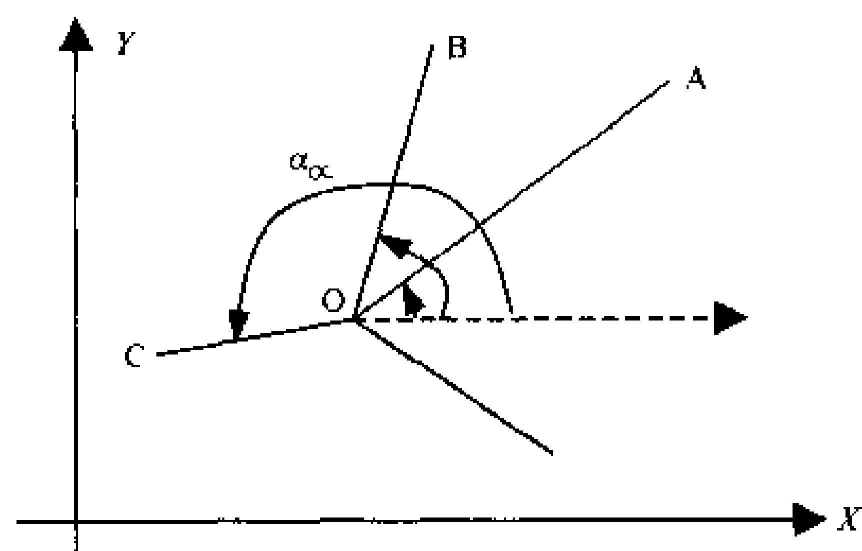


图 14-6　同一结点弧段的计算方位角

14.2.2　基于 Q_i 的算法

Q_i 算法(齐华等,1996)的基本思想是将同一结点出发的射线之间的关系转换为射线与结点外围正方形交点的相邻关系。

如图 14-7 所示,关于结点 O(x_o, y_o)对称的外围正方形为 ABCD,定义 ABCD 与 Y 轴正向的交点 N 为长度计算的参照点。从结点 O 出发的任意射线与该正方形有且仅有一个交点,定义交点按逆时针方向沿正方形边界行走到 N 点的长度 L 为算法参数。若射线 OS 与正方形 ABCD 的交点为 P,且 ABCD 的边长定义为 2 个单位,则 P(x_p, y_p) 点到 N 点的长度 L_P 为:

$$L_P = L_{PC} + L_{CB} + L_{BA} + L_{AN}$$

由于:$L_{PC} = 1 - \frac{\Delta y}{\Delta x} = 1 - \frac{y_p - y_o}{x_p - x_o}$;

$L_{CB} = L_{BA} = 2$;

$L_{AN} = 1$;

故

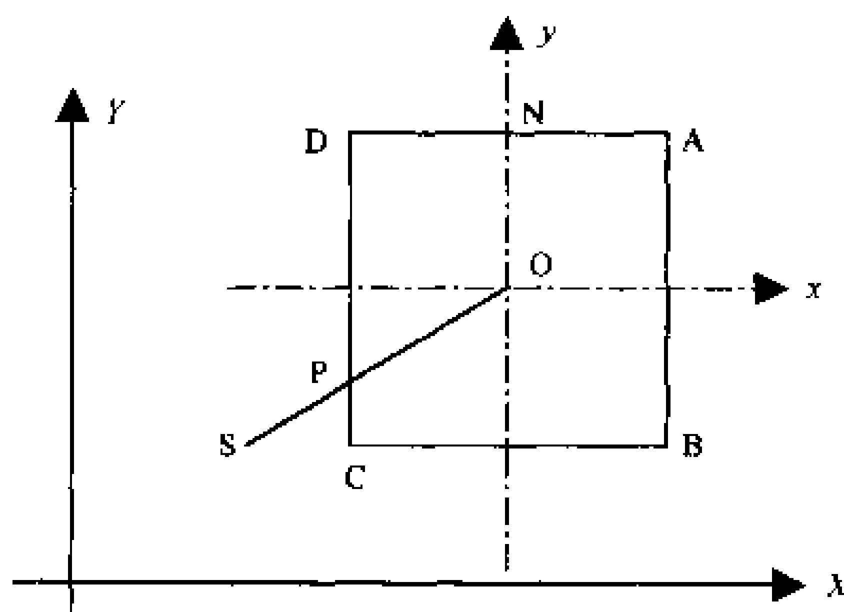


图 14-7　Q_i 算法原理(齐华等,1996)

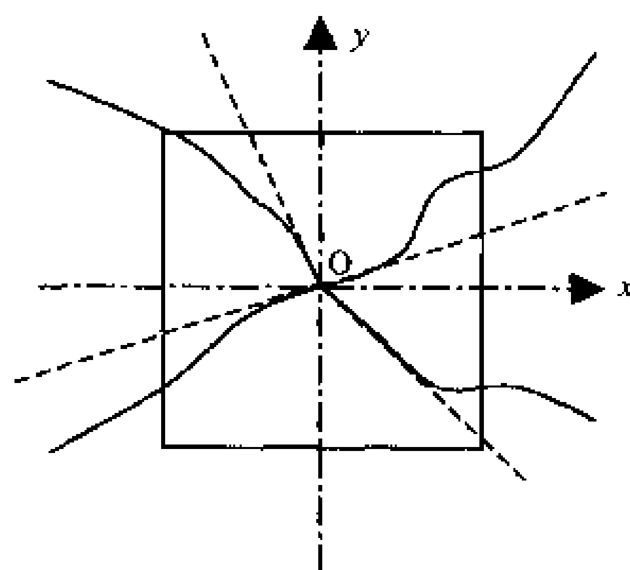


图 14-8　Q_i 算法实施策略(齐华等,1996)

$$L_P = 6 \quad \frac{y_p - y_o}{x_p - x_o}$$

齐华等(1996)给出了计算任意方向射线与外围正方形的交点 i 到 N 点的折线长度 L_i 的分段函数 $Q_i(x, y)$：

$$Q_i(x,y) = \begin{cases} \Delta x_i/\Delta y_i \cdots & \text{当}(\Delta y_i > 0) \cup (\Delta x_i > 0) \cup (\Delta y_i > \Delta x_i) \\ 2 - \Delta y_i/\Delta x_i \cdots & \text{当}(\Delta y_i > 0) \cup (\Delta x_i > 0) \cup (\Delta y_i \leqslant \Delta x_i) \\ 2 - \Delta y_i/\Delta x_i \cdots & \text{当}(\Delta y_i \leqslant 0) \cup (\Delta x_i > 0) \cup (\Delta x_i > -\Delta y_i) \\ 4 + \Delta x_i/\Delta y_i \cdots & \text{当}(\Delta y_i < 0) \cup (\Delta x_i > 0) \cup (\Delta x_i \leqslant -\Delta y_i) \\ 4 + \Delta x_i/\Delta y_i \cdots & \text{当}(\Delta y_i < 0) \cup (\Delta x_i \leqslant 0) \cup (\Delta y_i \leqslant \Delta x_i) \\ 6 - \Delta y_i/\Delta x_i \cdots & \text{当}(\Delta y_i \leqslant 0) \cup (\Delta x_i < 0) \cup (\Delta y_i > \Delta x_i) \\ 6 - \Delta y_i/\Delta x_i \cdots & \text{当}(\Delta y_i > 0) \cup (\Delta x_i < 0) \cup (\Delta y_i \leqslant -\Delta x_i) \\ 8 + \Delta x_i/\Delta y_i \cdots & \text{当}(\Delta y_i > 0) \cup (\Delta x_i \leqslant 0) \cup (\Delta y_i > -\Delta x_i) \end{cases}$$

(14-6)

算法具体实施时，对于同一结点的所有弧段，各选择一个最靠近结点的另一采样点；以结点为始点，以采样点为连接点，构成射线，如图 14-8 所示。计算各射线的 Q_i 值，按 Q_i 进行排序，则得到相应弧段的空间邻近关系。

14.2.3 基于矢量外积的算法

基于角度的算法和基于 Q_i 的算法均是通过计算弧段某一参数的绝对值来完成弧段的排序，从而得到弧段之间的相对位置关系。高云琼等(2002)基于计算几何矢量外积的思想，提出可以直接通过对弧段之间的矢量外积的比较来确定弧段之间的相对位置关系，而不必进行弧段绝对位置的计算。

1. 同一结点两弧位置关系判断

设从结点 $O(x_o, y_o)$ 出发的 n 条弧段序列为 $V = \{P_1, P_2, \cdots, P_n\}$。若弧段序列 V 中任意两条弧段 P_s、P_t 的另一端点分别为 $P_s(x_s, y_s)$ 和 $P_t(x_t, y_t)$，则该两弧相对于结点 O 的矢量外积为：

$$P_s \times P_t = \begin{vmatrix} x_s - x_o & x_t - x_o \\ y_s - y_o & y_t - y_o \end{vmatrix} = (x_s - x_o)(y_t - y_o) - (x_t - x_o)(y_s - y_o)$$

(14-7)

当 $(P_s \times P_t) > 0$ 时，相对于结点 O 来说，弧段 P_s 位于弧段 P_t 的顺时针方向；当 $(P_s \times P_t) < 0$ 时，相对于结点 O 来说，弧段 P_s 位于弧段 P_t 的逆时针方向；当 $(P_s \times P_t) = 0$ 时，出现边界条件，此时弧段 P_s 与 P_t 共线，方向相同或相反。

2. 同一结点任意弧段排序

排序过程为：① 从一结点出发的原始弧段序列中，以某种方式选择一条弧段作为基准弧段；② 将位于基准弧段逆时针方向上的弧段置于弧段序列中该弧段之前，而位于基

准弧段顺时针方向上的弧段则置于序列中该弧段之后;③ 当边界条件出现时,表明该弧段与基准弧段共线异向,将其置于基准弧段之前,然后利用某种排序方法通过矢量外积的比较完成局部排序。

高云琼等(2002)在试验过程中发现,利用矢量外积对弧段进行排序的过程同二叉树的构建过程十分吻合。因而,可以逐步插入法构建二叉树为例,来说明弧段排序过程如下:

设有从结点 $O(x_o, y_o)$ 出发的原始弧段序列 $V = \{P_1, P_2, \cdots, P_n\}$。若从 P_1 开始,每次选择 V 中未选取的某一弧段 P_i,按下列原则建立二叉树 Q 的一个新结点:① 若 Q 为空,则新弧段为 Q 的根结点弧段 P_o;② 若 Q 非空,则 P_i 与 P_o 进行矢量外积比较,设 $S = P_i \times P_o$,若 $P_i \times P_o \geqslant 0$,则将 P_i 插入 P_o 的左子树中;若 $P_i \times P_o < 0$,则将 P_i 插入 P_o 的右子树中;循环执行,直到 V 中的所有弧段均已选取并完成插入。在左、右子树中,再分别循环执行以上过程,完成树的细化。

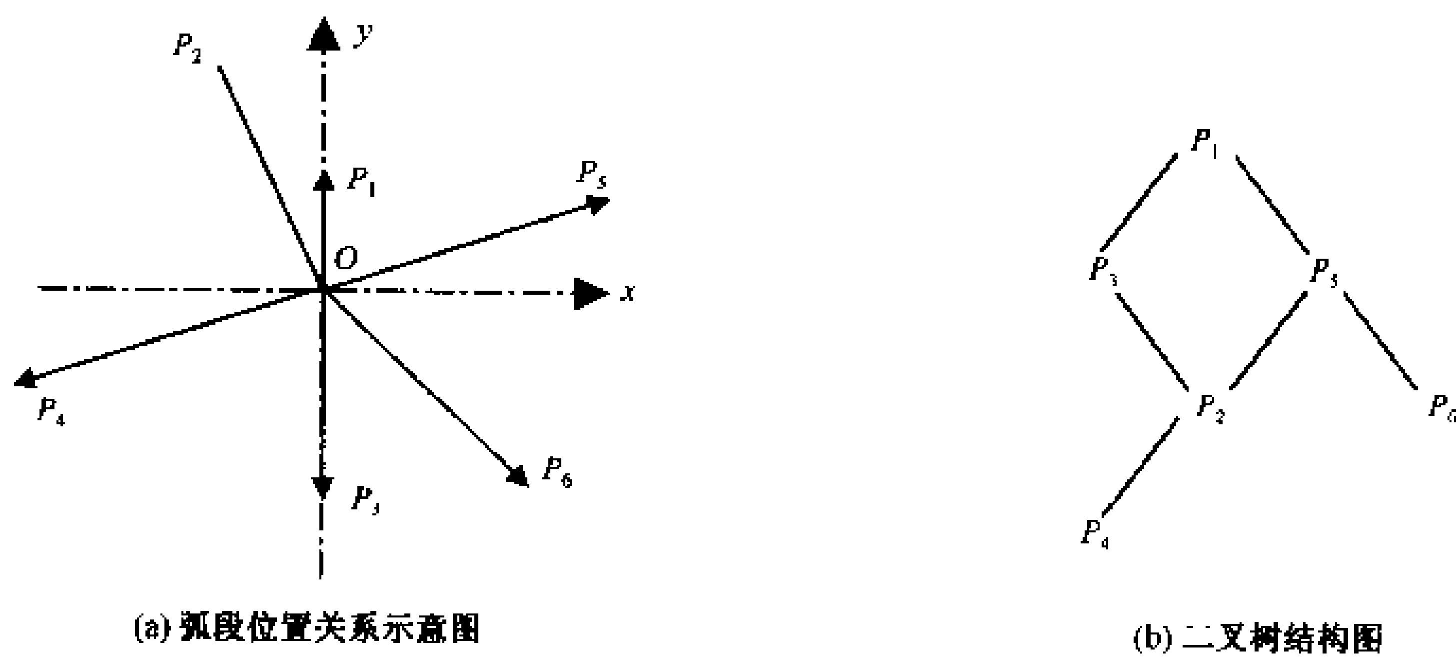


图 14-9 矢量外积算法排序实例

如图 14-9a 所示,从结点 O 出发的 6 条弧段为 1、2、3、4、5、6。按上述步骤执行排序过程,弧段提取顺序为 1、3、2、5、4、6,所生成的二叉树结构如图 14-9b 所示。二叉树排序完成后,对其进行中序遍历,即得到排好序的弧段序列,例中为 $V' = \{3, 4, 2, 1, 5, 6\}$。

对以上 3 种方法的时间效率进行比较,结果见表 14-1。

表 14-1 弧-弧排序算法的效率比较

基本操作	角度算法	Q_i 算法	矢量外积算法
	执行次数		
加、减法	2	2	1
乘、除法	1	1	1
Abs(x)	1	0	0
$\mathrm{Tan}^{-1}(x)$	1	0	0
判断	3	3	1
时间复杂度	$O(n)$	$O(n)$	$O[\log(n)]$

§14.3 多边形自动生成算法

多边形自动生成是 GIS 中面域数据组织和拓扑关系建立的重要环节。通常，可以基于方位角而采用左转算法或右转算法来生成多边形，即根据已经输入的结点和弧段信息，按规定的方向（若按顺时针方向，则称为右转算法；若按逆时针方向，则称左转算法），由系统通过弧段文件自动产生多边形信息，进而自动建立多边形文件。近年，人们针对左转算法提出了一些优化措施，如杨海宏等（1998）和闫浩文等（2000）；有人则提出了一些新的生成算法，如周立新等（1999）。

14.3.1 基于方位角的左转算法

多边形生成左转算法的基本思想是：从结点-弧段拓扑关系表（表 3-6）中的第一结点对应的第一弧段开始，沿逆时针方向搜索它所对应的多边形，通过对该弧段下一结点所对应的其它弧段的计算方位角的判断，确定该多边形的下一后继弧段；再以该后继弧段的下一结点判断其后继弧段，直到回到起始结点。然后跳转该表第一结点所对应的下一弧段，重新开始搜索另一多边形，直到第一结点所对应的弧段全部搜索完毕。再转入该表的下一结点，按上述规则重新开始。依此，直到生成了完整而不重复的多边形-弧段拓扑关系（表 3-4）和弧段-多边形拓扑关系（表 3-7）。

下面以图 14-10 为例，说明多边形-弧段拓扑关系自动生成的步骤与搜索结果：

1）依次计算各结点关联弧段的计算方位角（式 14-5），并从小到大排序，写入方位角排序表（数组）；

2）根据结点-弧段拓扑关系表（表 3-6）和点位坐标文件（表 3-2）读取首点（A）的点位，以此为追踪起始点；

3）根据方位角排序表查找首点（A）的第一弧段 AB，作为起始追踪边；

4）从方位角排序表中读取从 B 点出发的各弧段的计算方位角，令最小为 minBC；最大为 maxBC

5）比较：BA 弧的方位角与 maxBC、minBC 的大小；

a）若 minBC<<=maxBC（如 $ABDB$' 多边形所对应的弧段 BD），则队列中比稍小的 BCi 边（即 BD 边）为该多边形的后继弧段；

b）否则（如 $ABDB$' 多边形中结点 B' 所对应的弧段 B' D），则 maxBC（如 $ABDB$' 多边形中结点 B' 所对应的弧段 B' A）所对应的弧段 BCi 为该多边形的后继弧段；

6）写该多边形的多边形-弧段拓扑关系文件，将 BCi 排在 AB 之后；

7）记 $B=Ci$，并以此为追踪起始点；转第 4～第 6 步，然后转第 8 步；

8）判断 $B=A$?

a）若 B /= A；转第 7 步，继续；

b）若 $B=A$；终止，完成该多边形的多边形-弧段拓扑记录。

9）根据方位角排序表查找首点（A）的下一弧段 AB'，作为起始追踪边；生成与

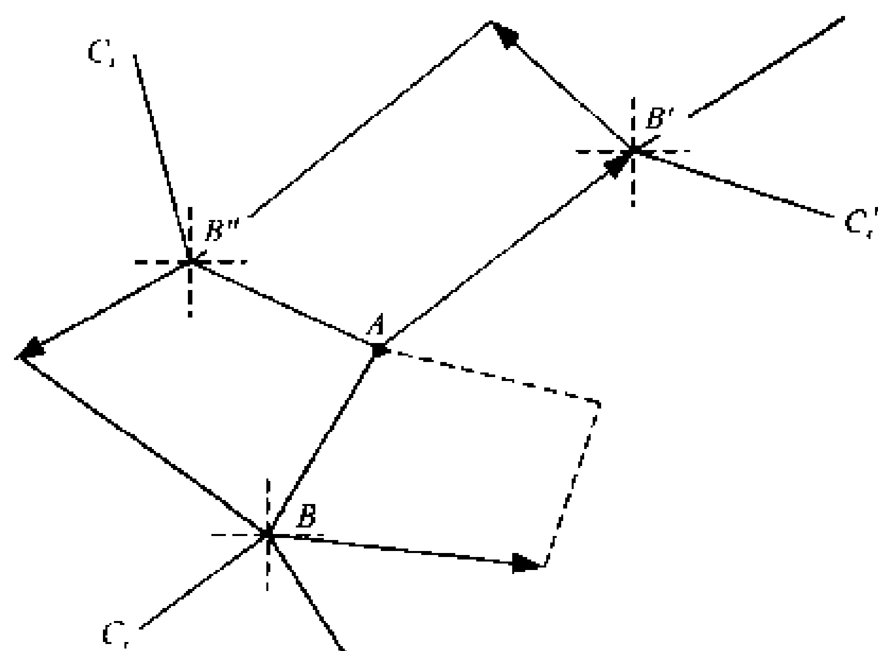


图 14-10　多边形生成左转算法实例

A 点关联的下一多边形……

10）判读首点（A）的所有弧段是否搜索完毕，若是，跳转结点－弧段拓扑关系表（表 3-6）的下一结点，转第 3）步；否则，转第 9）步。

11）如此循环，直到生成了完整而不重复的多边形-弧段拓扑关系（表 3-4）。

需要指出的是，弧段方位角的计算实质上是计算弧段 BC 的起始射线的方位角。弧段的起始射线定义为从弧段的起点 B 出发、并通过离起点 B 最近的弧段内点的直线。此外，若多边形包含岛，则由上述弧段和沿顺时针方向的岛边界共同构成有岛多边形。

多边形生成右转算法的基本思想是：在已建立弧段-多边形拓扑关系（表 3-7）的基础上，根据弧文件提供的左右多边形信息，将凡是与当前多边形有关的信息全部提取出来；若当前多边形位于某条弧段的右侧时，该弧段的起止结点顺序不变；否则交换其起止结点。然后将与当前多边形有关的全部弧段，按结点首尾相连的顺序，连接起来形成闭合多边形；若其中还包含岛，则由上述弧段和沿逆时针方向的岛边界共同构成有岛多边形。进而可以求出多边形的面积和周长，并将结果写入最终生成的多边形文件。

14.3.2　结点依次搜索算法

周立新等（1999）将点、弧段、多边形组成的图形看成为一个连通图，研究提出了一种从弧-弧邻接关系（以弧-结点表、弧的邻接表表示）自动生成多边形及其拓扑关系的算法。该算法的基本原理是以多边形的最小性和弧段的连接次数为基础，从某一结点开始搜索，在搜索过程中动态形成同级结点和次级结点，并依次搜索同级结点、次级结点和上级结点。所谓多边形的最小性，是指构成多边形的相邻弧段之间的夹角总是最小的，即多边形内部不再有任何支链，具有最小性的多边形称为最小多边形。所谓弧段的连接次数，是指某一弧段为多边形所共有的次数，最多为 2 次。

1. 算法描述

算法的实施步骤为（周立新等，1999）：

- 第一步：任取一边界结点为始点，作为当前结点；
- 第二步：对当前结点进行扩展，即选取某一与其相连且在当前结点没有被扩展的结点，判断：

案例 1：若相连结点在其他结点处没有被扩展过，则作为当前结点的次级结点，然后检查其他与当前结点相连的结点：

```
IF 所有相连结点均已扩展
    THEN 转第五步；
ELSE 转第二步；
```

案例 2：若相连结点在其他结点处已被扩展过，则形成一个多边形，转第三步；

- 第三步：搜索形成最小多边形：从当前结点出发，置队列 Q1 为空，

```
LOOP:
  IF 有同级结点相连且相应弧段的连接次数小于 2
    THEN 沿相连弧段前进一步，并将相连结点 P 加入队列 Q1；
  ELSE IF 有次级结点相连且相应弧段的连接次数小于 2
          THEN 沿相连弧段前进一步，并将相连结点 P 加入队列 Q1；
        ELSE 向上一级结点（相应弧段的连接次数小于 2）回溯一步，并将相连结点 P 加入队列 Q1；
        ENDIF
    ENDIF
  将结点 P 与 Q1 中的每一已知结点（起始结点除外）进行比较：
    IF 结点 P 与 Q1 中的结点 P1 间存在相连的弧段，
      THEN 将 P 与 P1 间的结点从 Q1 中删除（以保证多边形的最小性）；
      ENDIF
    IF 结点 P 为起始结点
      THEN 搜索得到一个多边形，转第四步；
    ELSE 转向 LOOP；
    ENDIF
```

- 第四步：每得到一个多边形，则将组成它的弧段记录下来，并赋予多边形一个编号，同时将这些弧段的连接次数加 1；检查与当前结点相连的结点，IF 所有相连结点都已被扩展

```
  THEN 转第五步；
ELSE 转第二步；
ENDIF
```

- 第五步：当前结点扩展完毕，

```
IF 存在没有被扩展的同级结点
  THEN 选取某一同级结点，转第二步；
ELSE IF 存在没有被扩展的次级结点
        THEN 选取某一次级结点为当前结点，转第二步；
      ELSE 转第六步；
```

ENDIF

ENDIF

- 第六步：所有弧段的连接次数为 2，搜索结束。

2. 算法实例

算法实例（周立新等，1999），如图 14-11 所示，已知各弧段的顶点坐标、结点坐标、结点编码、弧段编码、多边形标识码，弧段的邻接表、弧-结点表如表 14-2、表 14-3 所示。

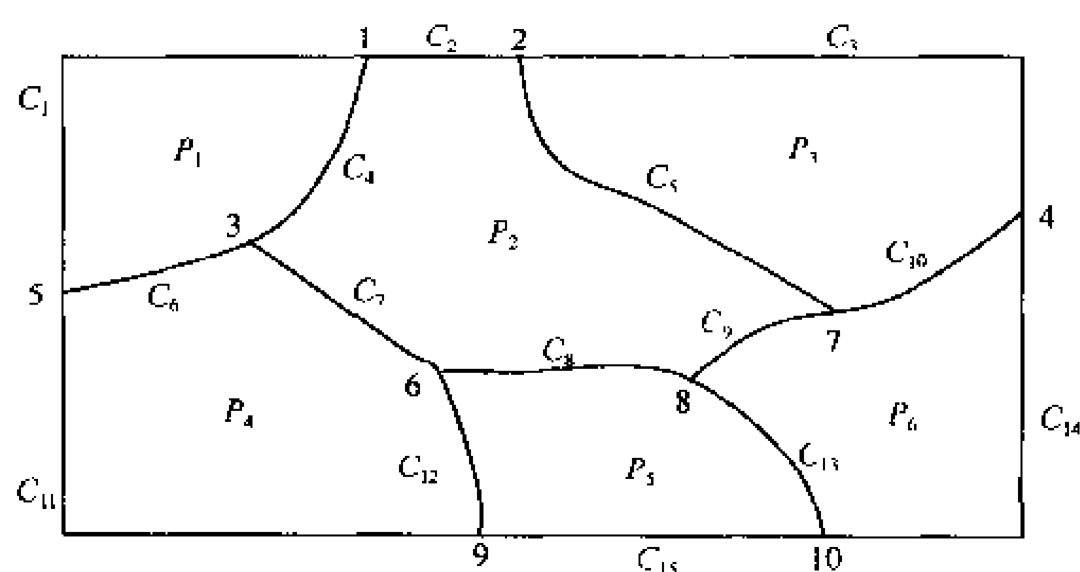


图 14-11 结点依次搜索算法

表 14-2 弧的邻接表

结点	相连的弧段
1	1，2，4
2	2，3，5
3	4，6，7
4	3，10，14
5	1，6，11
6	7，8，12
…	…

表 14-3 弧-结点表

弧号	起始结点	终止结点
1	1	5
2	1	2
3	2	4
4	1	3
5	2	7
6	3	5
…	…	…

其算法步骤：首先选取结点 1 为始点，作为当前结点进行扩展，依次扩展到结点 5、3、2，然后再选取结点 5 作为当前结点进行扩展……结点 1 为 1 级结点，5、3、2 为 2 级结点……扩展次序为结点 1、5、3、2、9、6、7、4、10、8，如图 14-12 所示。得到的多边形拓扑结构如表 14-4 所示。

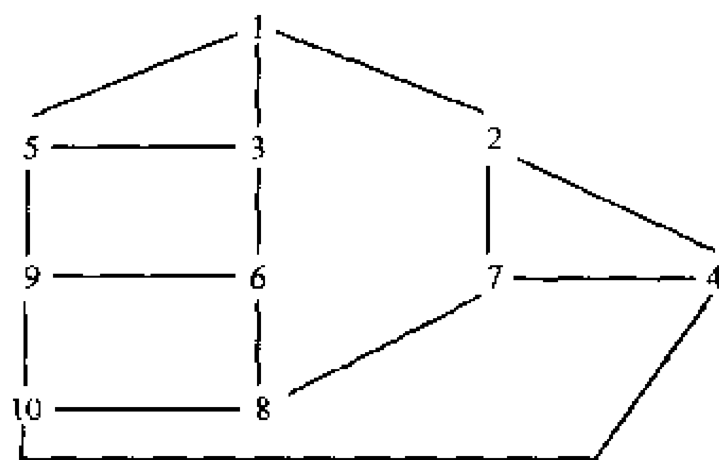


图 14-12 结点扩展顺序

表 14-4 多边形-弧段拓扑关系表

多边形号	相连的弧段
1	1，4，6
2	2，4，7，8，9，5
3	3，5，10
4	6，7，11，12
5	8，12，15，13
6	9，13，14，10

可以看出，该算法避免了进行多边形方位角的计算和比较，使用同级结点优先搜索和只搜索连接次数小于2的弧的策略，避免了对图中弧段邻接关系表的反复搜索，并减少了回溯深度，从而提高了速度。该算法的主要计算工作量在第2步，若某一结点有 m 条弧段相连（称该结点的度为 m），则该结点平均需要进行 $m(m+1)/2$ 次弧段搜索比较。而基于方位角的左转算法，该结点要计算 m 次反正切，并作 $m(m+1)/2$ 次角度比较。

§14.4 多边形图裁剪算法

裁剪是从数据集中提取信息的过程，多边形图的裁剪是计算机图形学的基本问题之一。多年来，人们围绕无拓扑多边形的裁剪，研究提出了许多算法（Mathew A. J.，1987；Andrew R. D.，1989；范慧琳，1997），涉及矩形窗口裁剪、任意窗口裁剪等。而在GIS中，面状地物多边形既有几何形状特征，又有拓扑特征。对这些具有拓扑关系的多边形的裁剪与计算机图形学中的纯粹几何形状的多边形（无拓扑关系）裁剪不同，GIS中裁剪算法必须维护多边形的拓扑关系，以保证裁剪前后多边形拓扑关系的正确性和继承性。

14.4.1 无拓扑多边形图裁剪算法

Weiler-Atherton算法是一种代表性的无拓扑多边形裁剪算法，主要适用于被裁剪的多边形与裁剪区域均为任意多边形的情形。该算法中的多边形用有序、有向的顶点环形表描述，当用裁剪区域来裁剪多边形时，裁剪多边形与被裁剪多边形边界相交的点成对出现，其一为入点，即被裁剪多边形进入裁剪多边形内部的交点；其二为出点，即被裁剪多边形离开裁剪多边形内部的交点。该算法的基本原理即为：由入点开始，沿被裁剪多边形追踪，当遇到出点时跳转至裁剪多边形继续追踪；如果再次遇到入点，则跳转回被裁剪多边形继续追踪。重复以上过程，直到回到起始入点，即完成一个多边形的追踪过程。

1. 算法步骤

设 PA 为被裁剪多边形，其顶点序列为 $A=\{A_0,A_1,\cdots,A_m\}(A_0=A_m)$；$PB$ 为裁剪多边形，其顶点序列为 $B=\{B_0,B_1,\cdots,B_n\}(B_0=B_n)$；用 PB 裁剪 PA 所得的多边形为 PC，其顶点序列为 $C=\{C_0,C_1,\cdots,C_s\}(C_0=C_s)$。三者的外边界顶点均按顺时针方向排列，内边界顶点均按逆时针顺序排列。裁剪算法步骤如下：

第一步：求 PA 与 PB 的边界交点，将交点（设为 $2K$ 个）分别加入 PA、PB 的顶点表中，新多边形记为 $PA'=\{A_0,A_1,\cdots,A_{m+2k}\}$，$PB'=\{B_0,B_1,\cdots,B_{m+2k}\}$；

第二步：建立交点表 $I=\{I_0,I_1,\cdots,I_{2k}\}$，记录交点类型及其在 PA、PB 顶点表中的位置；

第三步：在交点表 I 中取出一个入点 I_i，在 PA' 中找到 I_i 的位置并沿顺时针方向追踪 PA' 的顶点表，直到遇到下一个交点 I_j，将追踪得到的顶点序列加入 PC 中；

第四步：在 PB' 中找到 I_j 的位置，并沿顺时针方向追踪 PB' 的顶点表，直到遇到下一个交点，将追踪得到的顶点序列加入 PC 中；

第五步：跳转至 PA'，重复第三、第四步，直到回到起始交点，得到 PB 裁剪 PA 所得的内侧多边 PC。

若由出点出发，按上述步骤，反方向追踪则会得到外侧多边形 PC。

2. 算法实例

如图 14-13 所示（吴兵等，2000），$PA=\{A_1,A_2,A_3,A_4,A_5,A_1\}$，$PB=\{B_1,B_2,B_3,B_4,B_5,B_1\}$，$PA$ 与 PB 的交点集为 $I=\{I_1,I_2,I_3,I_4,I_5,I_6\}$。重构 PA、PB 多边形顶点序列得到：

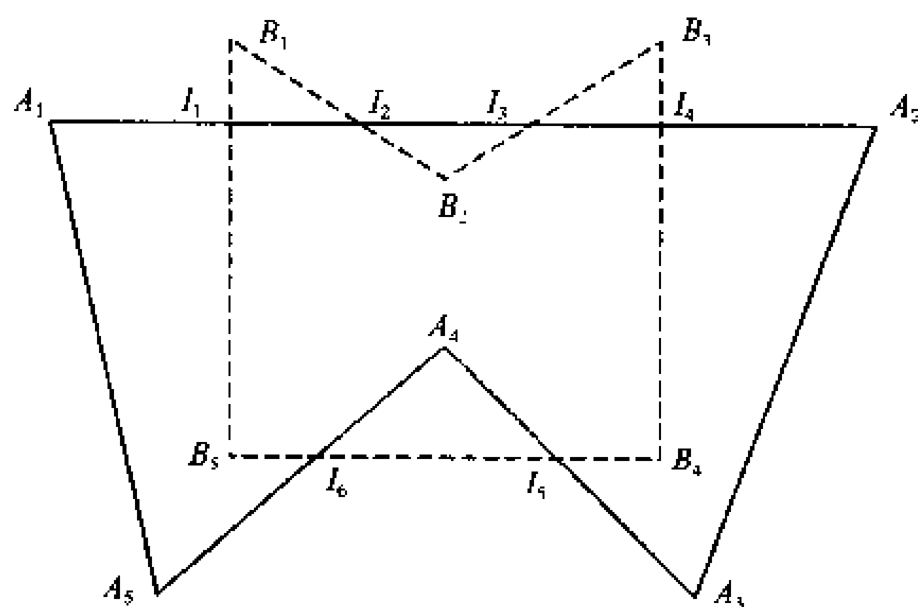


图 14-13 Weiler-Atherton 算法实例

$PA'=\{A_1,I_1,I_2,I_3,I_4,A_2,A_3,I_5,A_4,I_6,A_5,A_1\}$

$PB'=\{B_1,I_2,B_2,I_3,B_3,I_4,B_4,I_5,I_6,B_5,I_1,B_1\}$

PB 裁剪 PA 所得的内侧多边形为：

$PC_{内}=\{I_1,I_2,B_2,I_3,I_4,B_4,I_5,A_4,I_6,B_5,I_1\}$

PB 裁剪 PA 所得的外侧多边形为：

$PC_{外} = \{I_2, I_3, B_2, I_2\} \cup \{I_4, A_2, A_3, I_5, B_4, I_4\} \cup \{I_6, A_5, A_1, I_1, B_5, B_6\}$

图 14-14 所示为 Weiler-Atherton 算法的顶点追踪过程。

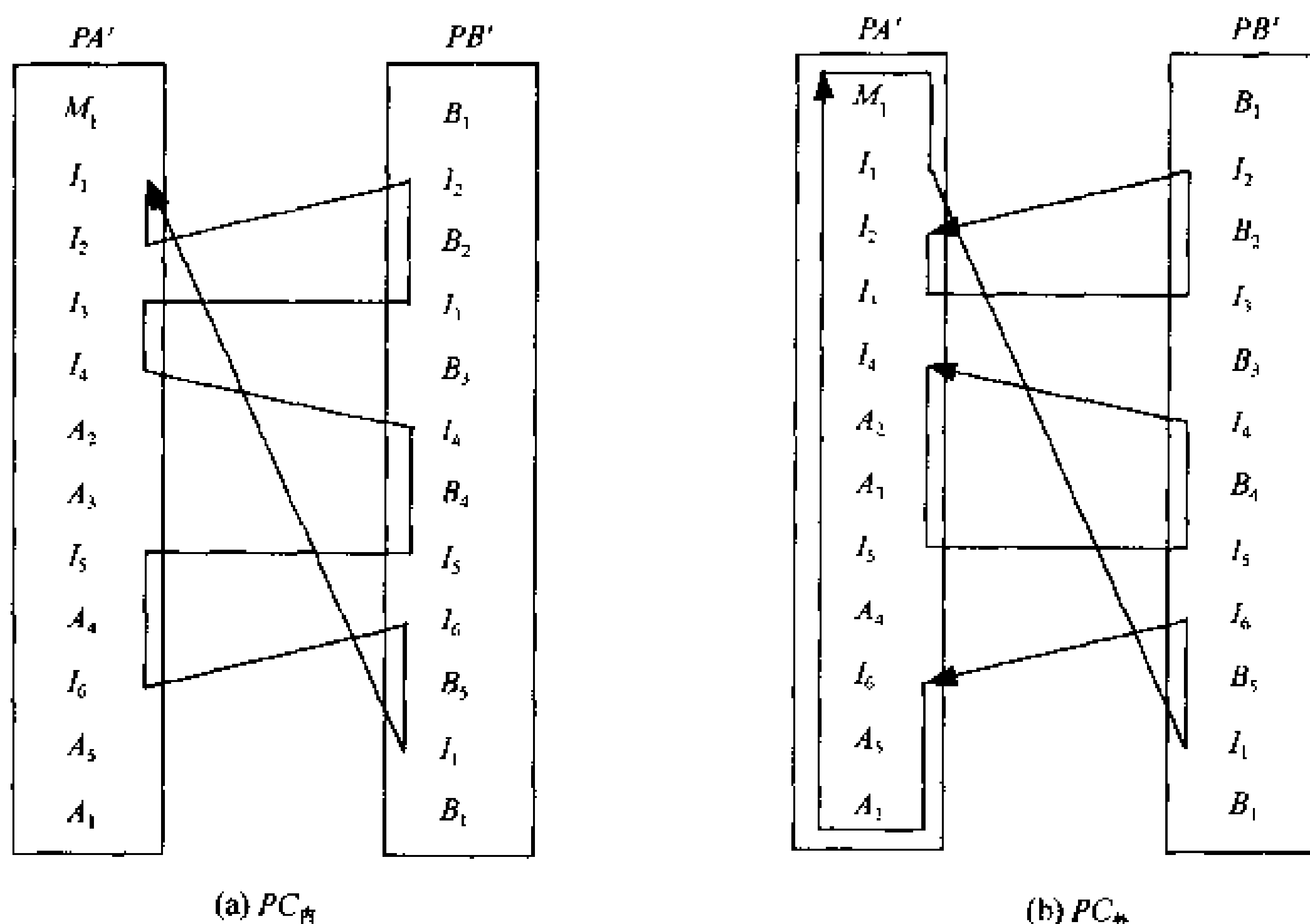


图 14-14　Weiler-Atherton 法的顶点追踪过程

14.4.2　有拓扑多边形图裁剪算法

上节介绍的 Weiler-Atherton 算法是以多边形的顶点序列为基础的，而具有拓扑关系的多边形并不是直接以顶点序列而是以弧段序列组成的。如果要套用 Weiler-Atherton 算法，必须将拓扑多边形分解为独立的多边形之后用顶点序列来描述，对单个多边形分别裁剪之后，再对裁剪结果重建拓扑关系。显然，这样得到的多边形拓扑关系与原多边形的拓扑关系之间不存在继承性，原拓扑关系及其他一些附加信息会丢失。

1. 新算法原理

鉴此，吴兵等（2000）提出按多边形弧段求交、多边形拓扑重组和追踪裁剪结果多边形 3 个基本步骤，将具有复杂拓扑关系的多边形裁剪问题转化为类 Weiler-Atherton 算法裁剪。

设区域 R 由一组具有空间拓扑关系的多边形组成，记为 $R = \{P_0, P_1, \cdots, P_n\}$；其中的任一多边形 P_i 均由一组有向弧段组成，记为 $P_i = \{A_0, A_1, \cdots, A_m\}$，$P_i$ 的外边界取 A_i 的顺时针方向，内边界取 A_i 的逆时针方向。令弧段由其顶点序列来描述，记 $A_i = \{V_0, V_1, \cdots, V_k\}$，其中 V_0 为起点，为 V_k 终点。除此之外，弧段与左右多边形的关系、结点与弧段之间的关系等均已知，即多边形的空间拓扑关系已经得到正确描述。

新算法增加了处理空间拓扑的步骤，用交点、弧段混合表取代原算法的交点表，将

原算法中追踪多边形顶点序列改造为追踪多边形弧段序列。从而保证当一个多边形被裁剪为多个多边形时，这些多边形会正确继承原多边形的拓扑信息及其附加属性，而不必在裁剪之后重建拓扑关系。

2. 新算法步骤（吴兵等，2000）

第一步：将 R 的所有弧段与裁剪多边形的弧段求交；

第二步：根据交点重组 R 的所有多边形与裁剪多边形，并维护原有拓扑关系；

第三步：对每个多边形建立交点、弧段混合表；

第四步：遍历所有多边形，反复执行第五至第八步；

第五步：从交点、弧段混合表中取出一个点，在被裁剪多边形中按弧段方向追踪，直到遇到下一个交点，将追踪所得的弧段序列加入到裁剪结果多边形中；

第六步：跳转至裁剪多边形相应位置。按弧段表方向追踪，直到遇到下一个交点，将追踪所得的弧段序列加入到裁剪结果多边形中；

第七步：跳转至被裁剪多边形相应位置，重复第五、第六步，直到回到起始交点处，完成一个多边形的追踪；

第八步：当多边形的交点、弧段缓冲表中的所有入点均追踪完毕，即完成此多边形的裁剪重构。

如图 14-15 所示，被裁剪区域 $R=\{P_1,P_2,P_3\}$，$P_1=\{A_1,A_4,A_5\}$，$P_2=\{A_2,-A_6,-A_4\}$，$P_3=\{A_3,-A_5,A_6\}$。

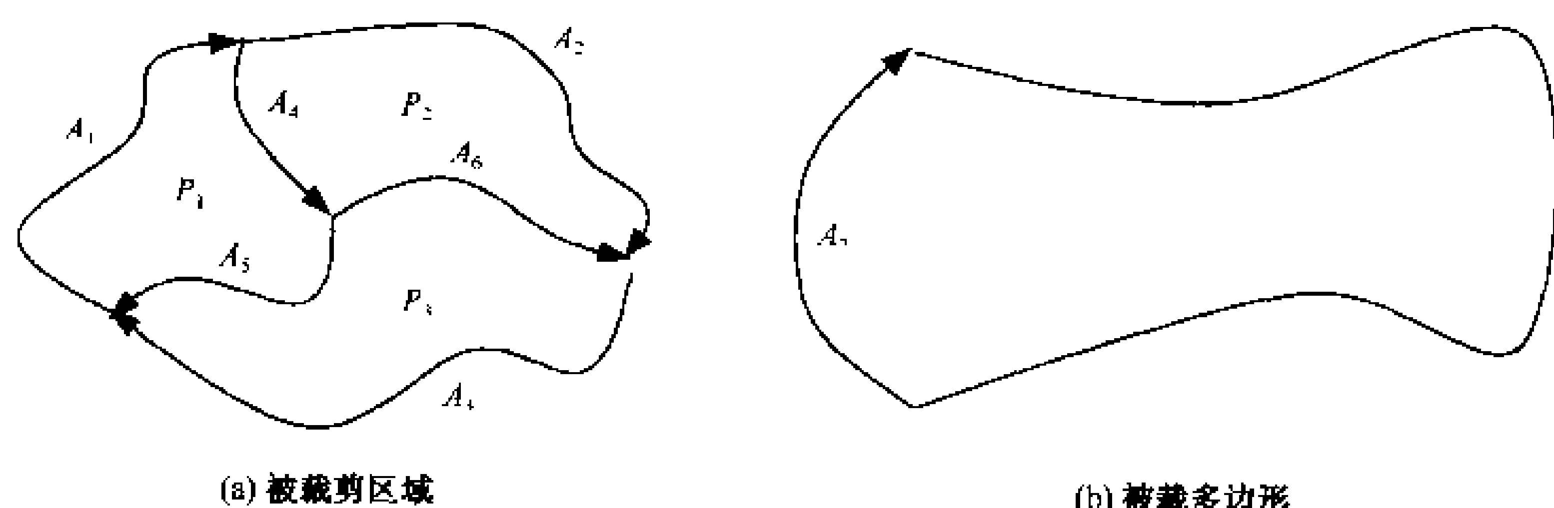


图 14-15　被裁剪区域与裁剪多边形示例

第一步：将裁剪多边形 $P_c=\{A_7\}$ 的弧段与 R 的所有弧段求交，得到交点集 $I=\{I_1,I_2,I_3,I_4,I_5\}$；

第二步：对 R 的所有多边形进行拓扑重组，得到：

$P_1=\{A_8,A_9,A_{15},A_{16},A_5\}$；

$P_2=\{A_{10},A_{11},-A_6,-A_{16},-A_{15}\}$；

$P_3=\{A_{12},A_{13},A_{14},-A_5,A_6\}$；

$P_c=\{A_{17},A_{18},A_{19},A_{20},A_{21}\}$；

第三步：以 P_1 为例建立多边形交点、弧段混合表：

$M_1=\{A_8,A_9,A_{15},A_{16},A_5\}$；

$P_c = \{I_1, A_{17}, I_2, A_{18}, I_3, A_{19}, I_4, A_{20}, I_5, A_{21}\}$；

在 M_1 中 I_2 为入点，由此点开始追踪内裁剪结果多边形，得到：

$P_{1内} = \{A_{16}, A_5, A_8, A_{17}\}$；

在 M_1 中 I_1 为出点，由此点开始追踪外裁剪结果多边形，得到：

$P_{1外} = \{A_9, A_{15}, -A_{17}\}$；

以同样的方法可得到其他两个多边形的裁剪结果：

$P_{2内} = \{A_{11}, -A_6, -A_{16}, A_{18}\}$；

$P_{3内} = \{A_{14}, -A_5, A_6, A_{12}, A_{20}\}$；

$P_{3外} = \{-A_{15}, A_{10}, -A_{18}\}$；

$P_{3外} = \{A_{13}, -A_{20}\}$；

最后，多边形 P_c 对区域 R 的拓扑裁剪结果为：

$P_内 = \{P_{1内}, P_{2内}, P_{3内}\}$；

$P_外 = \{P_{1外}, P_{2外}, P_{3外}\}$；

由图 14-16 可见，裁剪之前 P_1 与 P_2 拥有公共弧段 A_4，分别为 A_4 的右、左多边形。内裁剪之后，$P_{1内}$ 与 $P_{2内}$ 拥有公共弧段 A_{16}，分别为 A_{16} 的右、左多边形，即 $P_{1内}$ 与 $P_{2内}$ 在裁剪之后仍然是空间相邻关系，并且分别继承了 P_1 与 P_2 的各种属性信息。这表明经过本算法的拓扑裁剪之后的多边形的空间拓扑关系得以维持与继承。

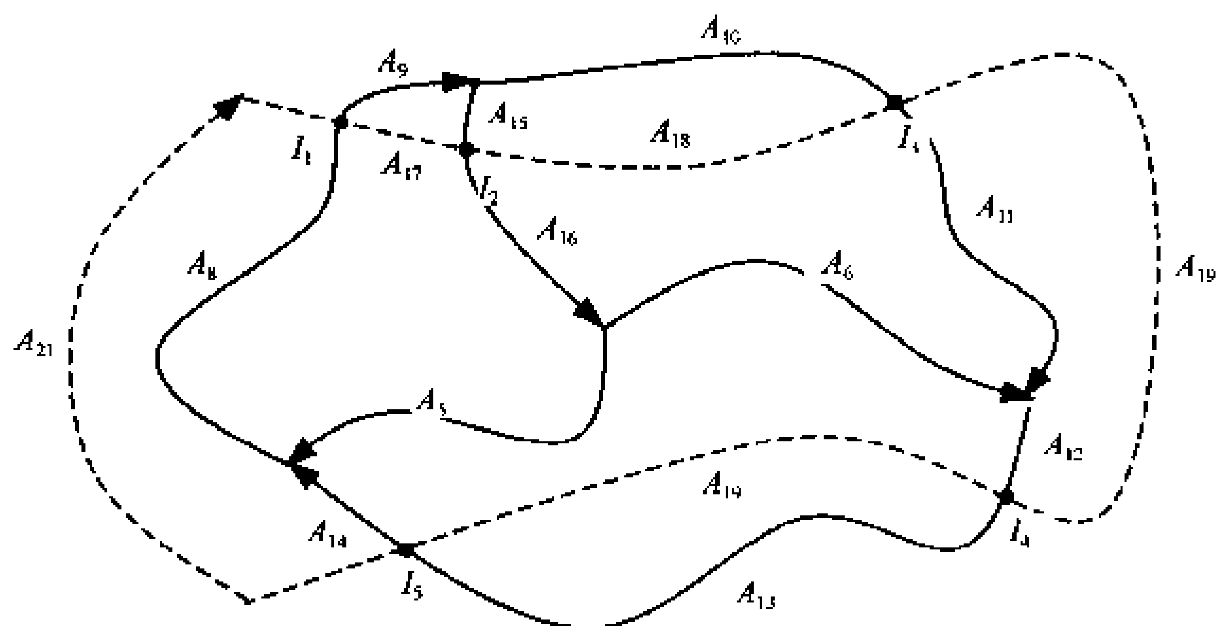


图 14-16 拓扑裁剪示例

参 考 文 献

钟振祥．1991．多边形内点的算法设计及其在 Geo-Union GIS 中的应用．三北防护林资源与环境动态监测信息系统研究．北京：测绘出版社．61～64

陈 春，张树文，徐桂芬．1996．GIS 中多边形图拓扑信息生成的数学基础．测绘学报．25(4)：266～271

齐 华．1997．自动建立多边形拓扑关系算法步骤的优化与改进．测绘学报．26(3)：254～260

齐 华，刘文熙．1996．建立结点弧-弧拓扑关系的 Qi 算法．测绘学报．25(3)：233～235

高云琼，徐建刚，唐文武．2002．同一结点上弧-弧拓扑关系生成的新算法．计算机应用研究．(4)：58～59

蔡少华，秦志远，朱 涛．1998．基于 MBR 的多边形内点自动生成算法．测绘工程．7(2)：61～63

闫浩文等．2000．基于方位角计算的拓扑多边形自动构建快速算法．中国图像图形学报．5A(7)：563～567

杨海宏等. 1998. 利用双邻点判断法优化拓扑关系自动生成算法. 中国图像图形学报. 3(7): 583-587
周立新,严 静,潘云鹤. 1999. 一个基于图的多边形拓扑关系生成算法. 计算机应用. 19(10): 37~39
范慧琳,杨 幸. 1997. 任意多边形区域的裁剪. 华侨大学学报(自然科学版). 18(4): 423~427
吴 兵,尹伟强,凌海滨. 2000. 具有拓扑关系的任意多边形裁剪算法. 小型微型计算机系统. 21(11): 166~168
Mathew A. J., 1987. Polygonal clipping of polylines. Computer Graphics Forum. 4(4): 407~414
Andreev R. D., 1989. Algorithm for clipping arbitrary polygons. Computer Graphics Forum. (8):183~191
Weiler K. and Atherton P., 1977. Hidden surface removal using polygon area sorting. Computer Graphics. 11: 214-222

第十五章　TIN的构建算法

二维平面域内任意离散点集的不规则三角网（triangular irregular network，TIN）是GIS数据表达、管理、集成和可视化的一项重要内容，也是地学分析、计算机视觉、表面目标重构、有限元分析、道路CAD技术等领域的一项重要的应用技术。TIN可以通过不同层次的分辨率来描述地形表面，也可以通过插入特征点、特征线、结构线等来精确逼近地表形态。

§15.1　Delaunay三角网

二维三角剖分的优化原则有多种，比较常用的有最大-最小距离原则、圆原则、最大-最小内角原则、最大-最小高原则、Thiessen原则等（姜寺山等，1995）。这些原则归结起来，无非是要求经剖分形成的三角网（TIN）要满足3点基本要求：① 惟一性：TIN是惟一的；② 最大最小角特性：三角形的最小内角尽量最大，即三角形尽量接近等边；③ 空圆特性：保证最邻近的点构成三角形，即三角形的边长之和尽量最小，且三角形的外接圆中不包含其他三角形的顶点（该特性也称Delaunay法则）。研究表明，在所有可能的TIN中，只有D-TIN（简称D-TIN）在满足上述要求时表现最为出色，D-TIN是给定区域点集的最佳三角剖分（Lee D. T. et al.，1980；Watson D. F. et al.，1981；Sapidis N. et al.，1991）。

15.1.1　Delaunay TIN基础

令$P=\{p_1, p_2, \cdots, p_n\}$为平面域（$R^2$）上的$n$个离散点的集合，有多种方法实现点集$P$的三角剖分。但是，俄国数学家Delaunay在1934年证明：必然存在且仅存在一种剖分算法，使得所有三角形的最小内角之和最大。后人将这种三角形称为Delaunay三角形。Delaunay三角形实质上是一种互相邻接且互不重叠的三角形的集合，其中的每一个三角形的外接圆均不包含其他三角形的顶点。

1. D-TIN基本概念

围绕Delaunay三角形还有一些重要概念，如形态比、角度特征向量、凸集、插入区域等。

（1）形态比：指三角形的内切圆半径与外接圆半径之比。TIN中所有三角形的形态比的平均值称为平均形态比。D-TIN是其所有TIN中平均形态比最大的。

（2）角度特征向量：将平面离散点集剖分出的所有三角形的角度按由小到大排队，所构成的向量。

（3）凸集（Convex）：每两点相互可视的平面点集。设S为平面上的点集，则：

① 如果 S 中的两点 A、B 可视，则 A、B 之直线连线位于 S 中；② 如果点 X 为 S 的观察点，则 X 与 S 中的每一点均可视；③ 凸集中每一点均是观察点。

（4）插入区域：外接圆包含待插入点的三角形的集合。

2．D-TIN 数据结构

以图 15-1 为例，其 D-TIN 的数据结构如表 15-1、表 15-2 所示。

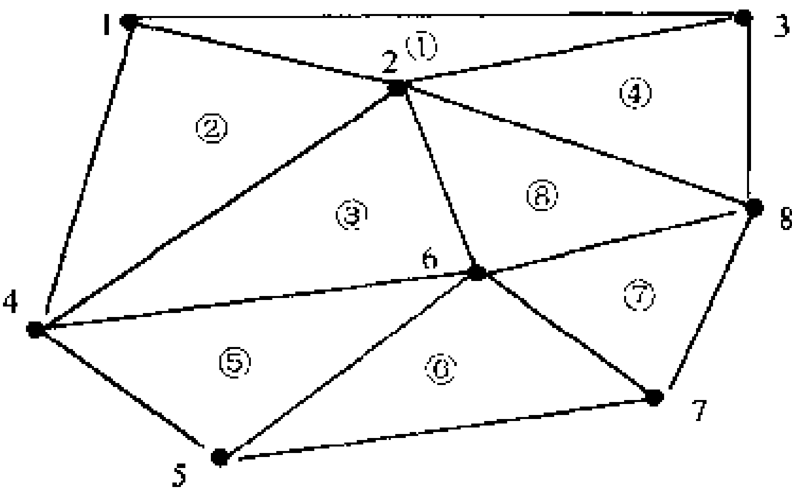


图 15-1 Delaunay 三角剖分实例

表 15-1 D-TIN 的离散点记录

标识号	x	y	z	最邻近关联点	关联三角形
1	50.00	100.00	200.00	3，2，4	①，②
…	…	…	…	…	…

表 15-2 D-TIN 的离散点记录

三角形标识号	顶点 1	顶点 2	顶点 3
①	1	2	3
…	…	…	…

3．D-TIN 局部优化

以图 15-2 为例，该图显示了在一个已存在的 Delaunay 三角形△123 中插入 4 号点，由于 4 号点位置的不同，根据 4 号点与△123 的关系和 Delaunay 法则，将产生不同的构 TIN 结果。其中图 15-2（c）所示为 4 号点位于△123 的外接圆上，两种构 TIN 结果都是有效的。但是，若以“三角网中所有边的长度的总和最小”为原则，则以短对角线的三角网为最佳选择。

Lawson（1977）提出根据“三角形最小内角最大化”法则来建立局部几何形状最优的 TIN：在由两个相邻三角形构成的凸四边形中，交换其对角线不会增加这两个三角形的 6 个内角的最小值。据此，Lawson 提出了局部优化方法（local optimization procedure，LOP）：通过交换凸四边形的对角线来获得等角性更好的 TIN。其应用实例如图 15-3 所示。

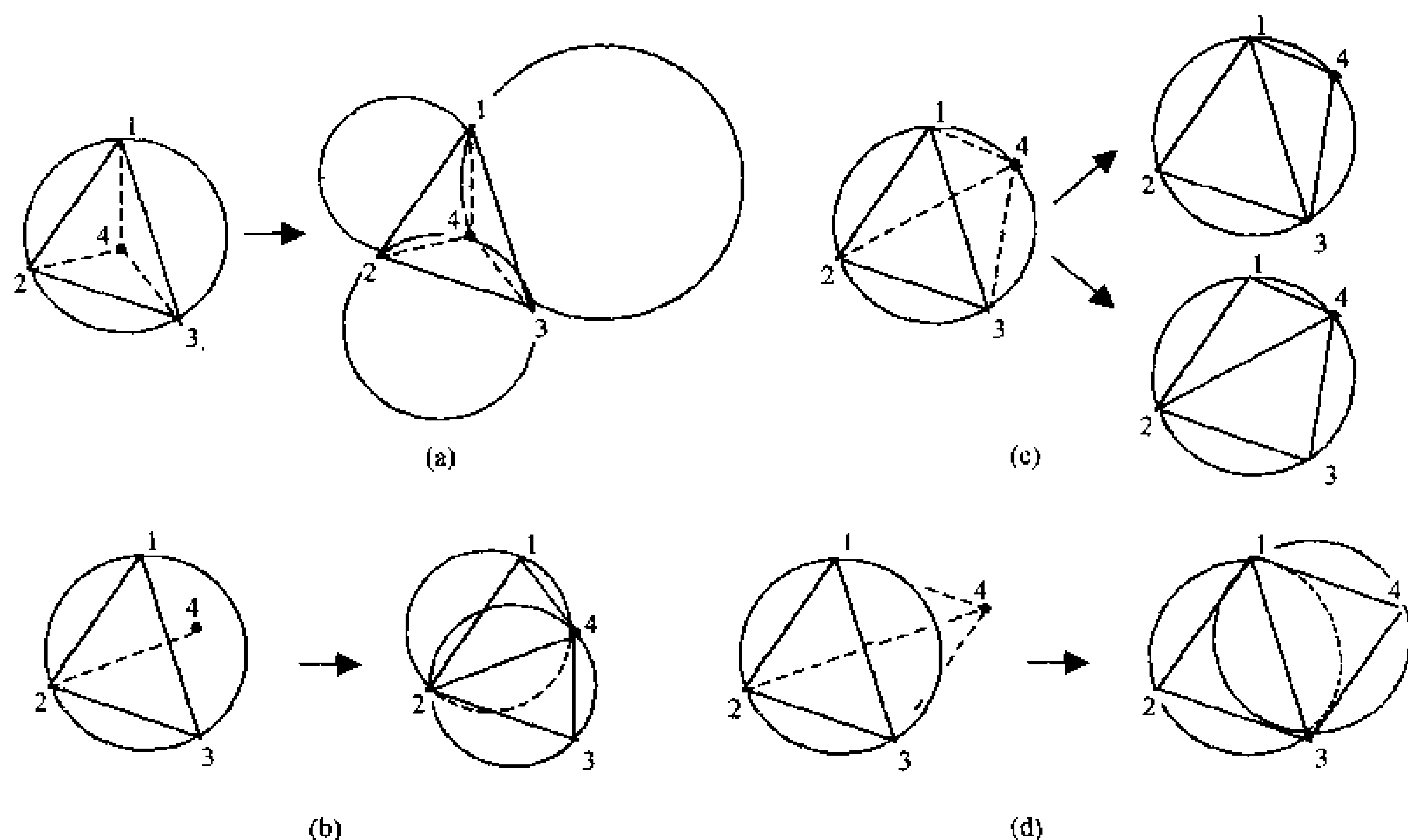


图 15-2　点与 D-TIN 之间的可能关系

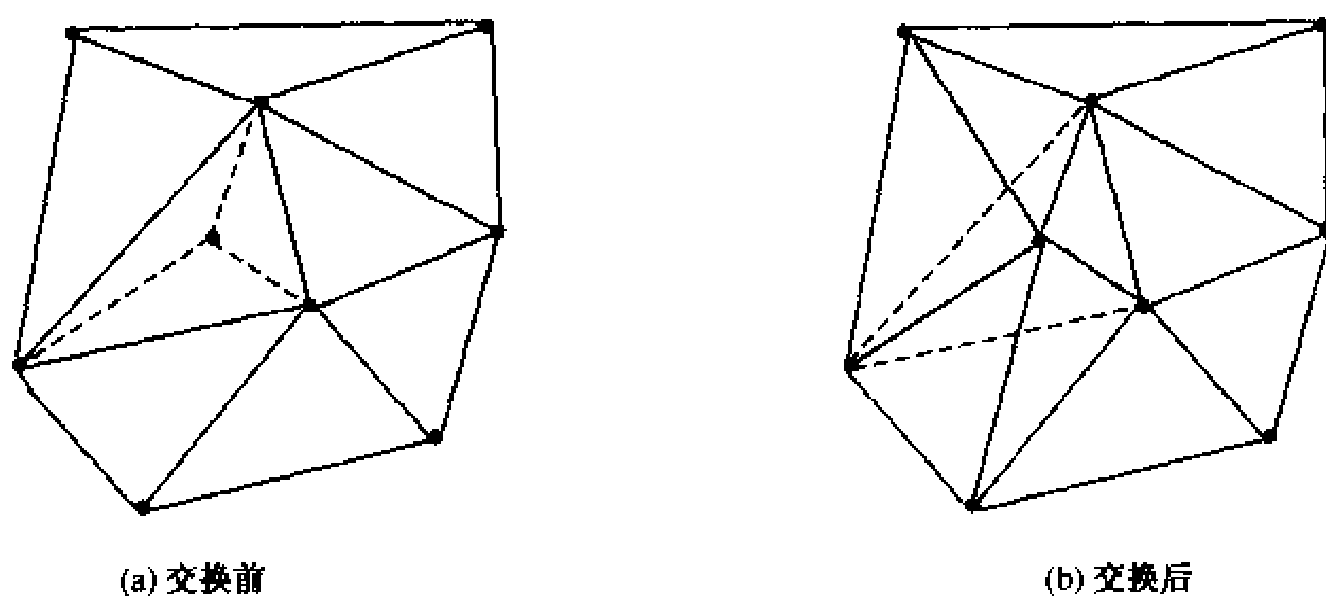


图 15-3　LOP 对角线交换示例

15.1.2　Delaunay TIN 构建算法

根据离散点域的数据分布特征和约束条件，三角剖分还可分为约束三角剖分和无约束三角剖分。在约束域下进行的 Delaunay 三角剖分称为约束 Delaunay 三角剖分（constrained delaunay triangulation 简称 CD-TIN）；在无约束域下的 Delaunay 三角剖分则称为常规 Delaunay 三角剖分（即 D-TIN）。D-TIN 的构建方法也可以分为间接法和直接法。所谓间接法，是通过计算平面点集的 Voronoi 图（见第 16 章）来得到对偶的 DT，通常采用这一方法；直接法包括基于 C^1 连续的复合三角 Bezier 曲面插值模型（柯映林，

1992）和基于最小生成树的算法（马小虎等，1997）等。

1. 基于最小生成树的 D-TIN 直接算法

以基于最小生成树的直接算法（马小虎，1997）为例，该算法的基本思想是：首先求出平面点集的欧几里德最小生成树（也称最小连通树）；然后逐次加入一条边，构成 TIN，直到最后一条边加入为止；最后按最大-最小内角的原则，通过局部变换得到 D-TIN。

设 $P=\{p_1, p_2, \cdots, p_n\}$ 为平面域（R^2）上的 n 个离散点的集合，连接 P 中所有点的总路径最短的树称为最小生成树（EMST（P））。设图 $G=(V, E)$，其中 V 是顶点集，E 是边集，若有 $e_{ik}=(v_i, v_k)$，$e_{ij}=(v_i, v_j)$，则称 e_{ik}与e_{ij}为一对相关边；有向边 v_iv_k 与 v_iv_j 之间的夹角称为相关边e_{ik}、e_{ij}的夹角。

定义由 EMST（P）或当前图中的相关边对组成的表为相关边对表，记为 IEPL（P）。在向 EMST（P）中加入边形成 TIN 时必须满足相交约束和包含约束条件：① 相交约束：三角形之间互不相交，即两个三角形除端点外无交点；② 包含约束：三角形之间互不包含，即任一个三角形不完全包含其他三角形。

算法的基本步骤为（马小虎等，1997）：

第一步：求出点集 P 的最小生成树 EMST（P）（见附 1）；

第二步：根据 EMST（P）求出初始相关边对表 IEPL（P）；

第三步：根据相关边对的夹角，从小到大排序相关边对表 IEPL（P）中的所有相关边对；

第四步：初始化 TIN 表 TML；

第五步：当 IEPL（P）中还存在一个边对没有被处理时，从中选择夹角最小的相关边对 e_{ik}、e_{ij}，并标志 mark 为 1；如果该相关边对构成三角形$\triangle v_iv_jv_k$，同时满足两个约束条件，则：① 把三角形$\triangle v_iv_jv_k$ 加入到 TML 中；② 把新边（v_i, v_k）加入到 EMST（P）中；③ 把新边（v_i, v_k）引入到新的相关边对，根据其夹角插入排序的 IEPL（P）中；

第六步：按最大-最小内角原则，通过局部优化变换（见附 2），得到 D-TIN。

1）*附 1：求 EMST（P）*

第一步：计算点集 P 中任意两点之间的距离，并存入距离数组 Dist［M］中。其中 $M=n(n-1)/2$，n 为P 的总点数，数组 Dist 中每一元素包含三项：Start 存放起点，End 存放终点，Length 存放两点之间的距离；

第二步：按 Dist［M］中元素的 Length 项的值从小到大排序数组 Dist［M］；

第三步：令 $k\leftarrow 1$，$e\leftarrow 0$，并置 EMST（P）空；

第四步：当 $e\leqslant n-1$ 时：①取 $u=$Dist［k］. Start，$v=$Dist［k］. End，并置 $k\leftarrow k+1$；②若边（u, v）与 T（P）中的边不构成回路，则把边（u, v）加入到 EMST（P）中，并置 $e\leftarrow e+1$。

2）*附 2：局部优化变换*

第一步：从 TML 表形成 TIN 的边表 EL；

第二步：对边表 EL 中的每一条非边界边进行操作：取出以该边为邻接边的两个相邻三角形构成一个四边形；如果该四边形是凹的，则不作任何处理；若是凸的，则计算两个三角形的最小内角 α_1；交换该四边形的对角线，重新计算两个新三角形的最小内角 α_2；如果 $\alpha_1 \geqslant \alpha_2$，则不作处理，否则；以新三角形代替原三角形，并按新三角形修改 TML，并令交换边计数变量 IE 加 1（初始值为零）；

第三步：如果 IE=0，则结束，否则转第一步。

2．CD-TIN 算法

目前，CD-TIN 构模方法主要有约束图法、分割-合并算法、加密点算法、三角形生长算法和分步法等。

1）约束图法

Lee 和 Lin 于 1986 年提出了基于约束图计算 CD-TIN 的构造算法。他们首先用 $O(n^2)$ 的时间计算约束数据域的可见性图，然后用 $O(n^2)$ 的时间检测优化可见性图并形成 CD-TIN。此方法从时间、空间的效率上都不够理想。

2）分割-合并算法

Lee 和 Schachter 在 1981 年提出了非约束数据域 Delaunay 三角剖分的分割-合并算法，Chew 将其推广至约束数据域，Joe 和 Wang 也提出了相类似的算法。分割-合并算法虽执行效率较高［$O(n\log n)$］，但约束线段的存在使得分割过程的实现比较困难，对于多约束条件下的三角网的生成，此方法存在很大不足。

3）加密点算法

Boissonnat 在进行 CD-TIN 构造的研究中，将约束数据转换为非约束数据，即对约束线段按空外接圆规则进行加密，然后对加密后的数据进行非约束数据域的 Delaunay 三角剖分。Boissonnat 认为加密过程的时间复杂度为［$O(n\log n)$］，虽然该算法简单易实现，但加密点的存在不但加大了数据量，而且改变了原来的数据集，前期数据处理较烦琐。

4）SHELL 三角化算法

该算法是 Piegl 和 Richard 在 1993 年提出的，其实质是非约束数据域的三角网生长算法的推广。算法由任一约束线段开始，以可见性和空外接圆特性为准则寻找可扩展的第三点。该算法关键是要解决好第三点的查询问题，以提高算法效率。

以上几种方法代表了当前约束三角网实现的主要研究成果，许多研究人员也提出其他的思路，但基本上是以上几种方法的变形。尽管 TIN 构模方法多种多样，根据数据来源的不同，可以将其分为基于离散点、基于等高线和基于平面点集的 Voronoi 图三种基本方式。其中基于平面点集的 Voronoi 图将在第 16 章中讨论。以下主要讨论基于离散点和基于等高线的构 TIN 算法。

§15.2 基于离散点的构 TIN 算法

15.2.1 无约束构 TIN 算法

基于离散点构 TIN 的基本思路是：利用随机分布的离散高程采样点建立连续覆盖整个区域的 TIN。其技术关键是确定哪 3 个离散数据点构成一个最佳三角形，并使得每个离散采样点均成为三角形的顶点。由于 Delaunay 三角形的良好特性，TIN 的构建一般均归结为 Delaunay 三角网（简称 D-TIN）。根据 D-TIN 的实现过程不同，可以分为逐点插入法、三角网生长法、分治算法、凸包法等。

1. 逐点插入法

Lawson (1977) 最早提出逐点插入法建立 D-TIN 的基本思想。此后，Lee，Watson，Floriani 和 Puppo 等众多学者先后完善和改进了这一算法。该法的基本思想是：在一个包含所有数据点的初始多边形中，将未处理的点逐次加入到已经存在的 D-TIN 中，每次插入一个点后，将 D-TIN 重新定义。如图 15-4 所示，该法的基本步骤是：

(1) 首先定义一个包含所有数据点的初始多边形；

(2) 从数据集中任意选择一个点（如 A），插入到初始多边形中建立初始三角网；

(3) 然后按以下步骤进行迭代计算，直至所有数据点均被处理：①插入一个离散采样点 P，在初始三角网中找出包含 P 的三角形 T，把 P 与 T 的三个顶点相连，生成 3 个新的小三角形；②用局部优化方法（local optimization procedure，LOP）从里到外优化三角网，确保所有点内插后的三角网为 D-TIN。

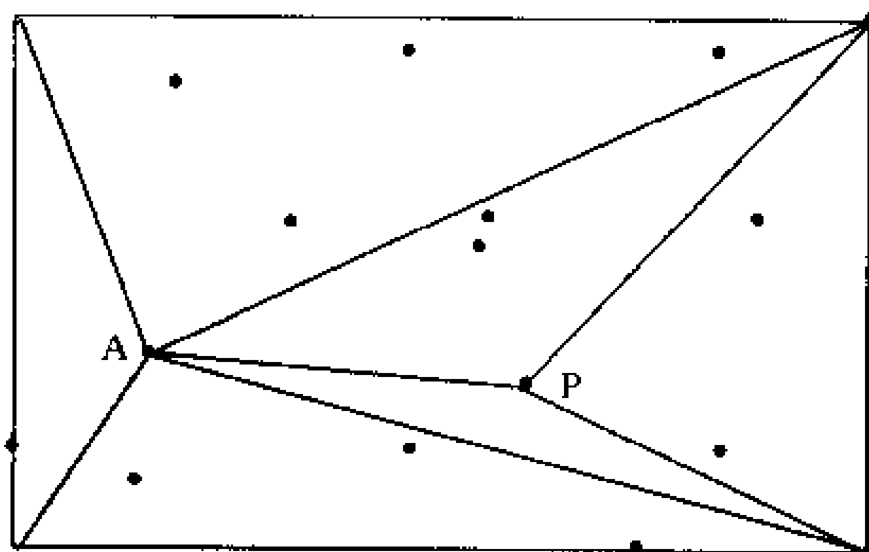


图 15-4 逐点插入法的原理

逐点插入法有多种不同的实现方式，其差别主要在于定义的初始多边形不同，如超三角形（super triangular）、矩形等。为便于处理凸多边形内的点，Macedonio 和 Pareaschi (1991) 对基于矩形的逐点插入法作了改进，假设点集内有 N 个点，其构 TIN 步骤为：

(1) 首先删除数据集中所有与矩形顶点相重合的点；

(2) 然后将矩形划分为 $\sqrt{N}$ 块或更小的矩形区域；

(3) 对各分块中的数据排序；

(4) 构建各分块的局部 D-TIN；

(5) 应用 LOP 法由里向外更新局部 D-TIN 的所有邻接三角形；

(6) 重复第（5）、（6）步，直到处理完所有点并生成全局 D-TIN。

2. 三角网生长法

三角网生长算法的基本思路是：首先找出离散点集中相距最短的两点，连线成为 D-TIN 的初始基线；然后按 D-TIN 的判断法则找出包含此基线的 Delaunay 三角形的第 3 端点，该端点应位于基线右侧；连接新点与原来两点形成两条新边；再按 D-TIN 的判断法则找出包含此两边的另外两个 Delaunay 三角形的第 3 端点；依次循环处理所有新生成的边，直到所有离散点均成为 D-TIN 的端点。也有人称这种三角网生成算法为“炸弹法”，其算法的基本步骤为：①变量与数组设计：设计两个计数变量 K 和 L，K 记录已扩展的三角形数，L 记录已形成的三角形数；在第一个三角形形成之后，分别置 $L=1$ 和 $K=1$。设置三个数组 $t_1[\,]$、$t_2[\,]$、$t_3[\,]$用于存放所构成的三角形的顶点编号。②确定首三角形：任选一个离散高程点作为首三角形的第一顶点；找出离该点最近的点作为首三角形的第二顶点；再寻找相距第 1、第 2 顶点之连线的中点最近、位于其右侧且不与两顶点共线的点作为首三角形的第 3 顶点；③扩展三角形：首先从 K 号三角形的第 1 条边 $(t_1[k]-t_2[k])$向外扩展。

(1) 寻找可能的扩展点：显然，以 $t_1[k]-t_2[k]$边为界，位于 $t_3[k]$同侧的离散点应排除在外，如图 15-5 所示。可以利用直线方程及其正负区的判别原理来实现这一过程。令 $t_1[k]$、$t_2[k]$的平面坐标分别为(x_1,y_1)、(x_2,y_2)，则 $t_1[k]-t_2[k]$边的直线方程为：

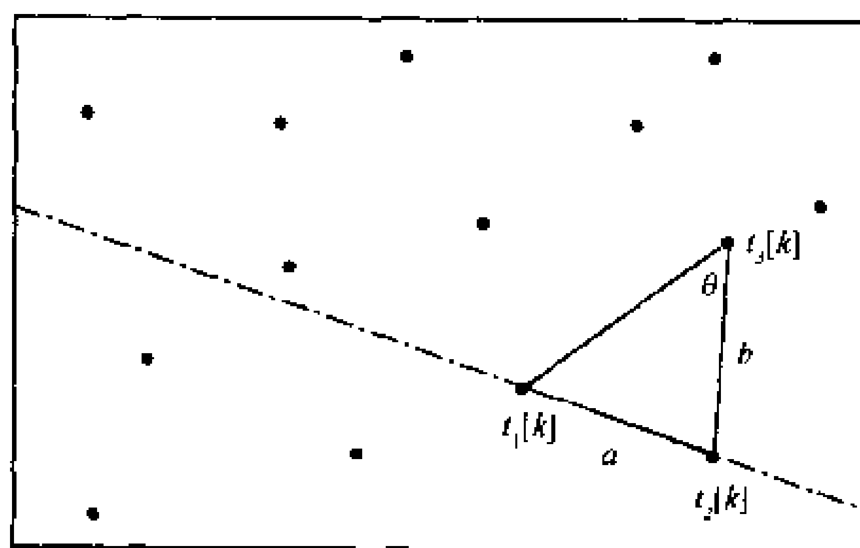


图 15-5 可能扩展点的直线判别法

$$f(x,y)=y-(ax-b) \tag{15-1}$$

式中：$a=\dfrac{y_2-y_1}{x_2-x_1}$；$b=\dfrac{x_1y_2-x_2y_1}{x_2-x_1}$。

若有离散点 P，其平面坐标为(x,y)；将 P 点及 $t_3[k]$点均代入上式，并判断：

$$\begin{cases} f(x,y)>0 & \text{P 点位于 } t_1[k]-t_2[k] \text{ 边的正区} \\ f(x,y)=0 & \text{P 点位于 } t_1[k]-t_2[k] \text{ 直线上} \\ f(x,y)<0 & \text{P 点位于 } t_1[k]-t_2[k] \text{ 边的负区} \end{cases}$$

若 P 点及 $t_3[k]$的结果同号(即同为正区或负区),则 P 点不是待扩展点;若异号(即互为正区或负区),则 P 点是可能的扩展点。

(2) 判断扩展点:利用三角形余弦定理,在所有可能的扩展点中,找出扩展边张角最大的点,该点即为满足 Delaunay 三角形条件的扩展点,将该点作为扩展三角形的第 3 个顶点,将其坐标送入 $t_3[L]$中。新三角形基本形成后,令 $L = L + 1$。

(3) 检查:为避免重复和交叉,要检查新三角形的 3 条边中是否有某条边被已形成的三角形用过两次:若是,则该三角形无效,令 $L = L - 1$;若非,则该三角形被确认。

(4) 第一条边($t_1[k] - t_2[k]$)向外扩展完后,再采用同样方法扩展 K 号三角形的第 2、第 3 条边。之后,令 $K = K + 1$,转下一个三角形继续扩展。当且仅当 $K = L$ 时,三角网构建完毕。

这种算法的大部分时间均花费在搜索符合要求的给定基线的邻域点过程中。也可以通过计算初始基线的外接圆的圆心和半径来完成对领域的搜索。为了节约搜索时间,McCullagh 和 Ross 给出了一个先将数据分块并排列成有序结构的改进算法。

此外, Mirante 和 Weingarten 还提出了一个径向扫描(Radial Sweep)算法,从随机分布的离散点集的外部边界开始,由外向内建立 TIN,其实施过程如下:

(1) 查找离数据重心最近的点为起始点;

(2) 计算起始点与其他所有点的距离和方位;

(3) 根据方位、距离将所有其他点按升序排序;

(4) 径向扫描各数据点,将起始点与各径向点相连,并连接相邻线段的任意两连续端点,生成辐射三角形;对于具有相同方向的点,则生成具有公共边的三角形对;

(5) 将辐射状三角形边界上的点连接,填充边界上的凹孔,以此方式由里向外生成新的三角形,直到达到数据的边界;

(6) 通过检查由一对相邻三角形形成的四边形两对角线的长度(类似于 LOP),从外部边界上的三角形开始由外向里顺序优化 TIN。

3. 分治算法

Shamos 和 Hoery 基于 Voronoi 图的构建,提出了分治算法的思想。Lewis 和 Robinson(1978)则将分治算法的基本思想用于 D-TIN 的构建。他们采用递归分割点集直至每个子集中仅含 3 个离散点而形成三角形的办法,经过自下而上逐级合并生成最终的三角网。此后, Lee 和 Schachter(1980)又改进和完善了 Lewis 和 Robinson 的算法,首先将数据排序,分成两个互不相交的子集,在每一个子集中建立 TIN 后,将两个 TIN 合并以生成最终的 DelaunayTIN;Dwyer(1987)对分治法作了进一步的改进,通过将数据分成垂直条块,进而用相交边界将条块再一次细分为区域,并应用带约束条件的 Lawson LOP 将对角线进行交换,因而可以处理带约束条件的数据。不同的实现方法主要区别在于点集划分、子三角形网生成与合并方法的不同。

对于给定的 n 个互不重合平面点中的数据集 V,分治算法的基本步骤为:

(1) 以横坐标为主、纵坐标为辅,将离散点集 V 按升序排列,即 $(x_i, y_i) < (x_{i+1}, y_{i+1})$,其成立条件为:$x_i \leqslant x_{i+1}$, $y_i < y_{i+1}$;

(2) 将离散点集 V 分成近似相等的两个子集 V_L 和 V_R,其中 V_L 包括数据中的前

一半，V_R 包括其余部分；

（3）在子集 V_L 和 V_R 中分别生成 D-TIN；并用 LOP 局部优化算法优化所生成的三角网；

（4）计算 V_L 和 V_R 的凸包，找出连接 V_L 和 V_R 的公共边及其邻接的两个三角形；

（5）从凸包下部的公共边开始，由底往上合并其邻接三角形。如果某一公共边的一个端点位于其中某一三角形的外界圆内，则将此非 Delaunay 边删除；然后应用 Delaunay 法则选择正确的 Delaunay 边；

（6）重复第（2）～（5）步，直到最终的 TIN 建立完毕。

总体上，分治算法以递归方式将数据排序成大小相同的单元子集，直至每一个单元子集能同样处理。一个规范的单元子集包括 3～6 个点，使用其中的前 3 个点生成一个三角形，然后将其余的点逐个插入即可建立各单元子集的局部 D-TIN。然后，从单元子集开始，将其局部 D-TIN 合并；依次往上进行合并，直到生成最终的全局 D-TIN。因此，分治算法的合并原则是：同层优先，由下往上，递归进行。

4. 凸包算法

所谓凸包（convex hull）是由包含二维平面上的点集中所有点的最小凸多边形。在凸包中，连接任意两点的线段必须完全位于多边形内。凸包是数据点的自然极限边界，相当于包围数据点的最短路径。显然，凸包必定是该点集的 D-TIN 的外部边界。以图 15-6 为例，凸包的产生过程为：

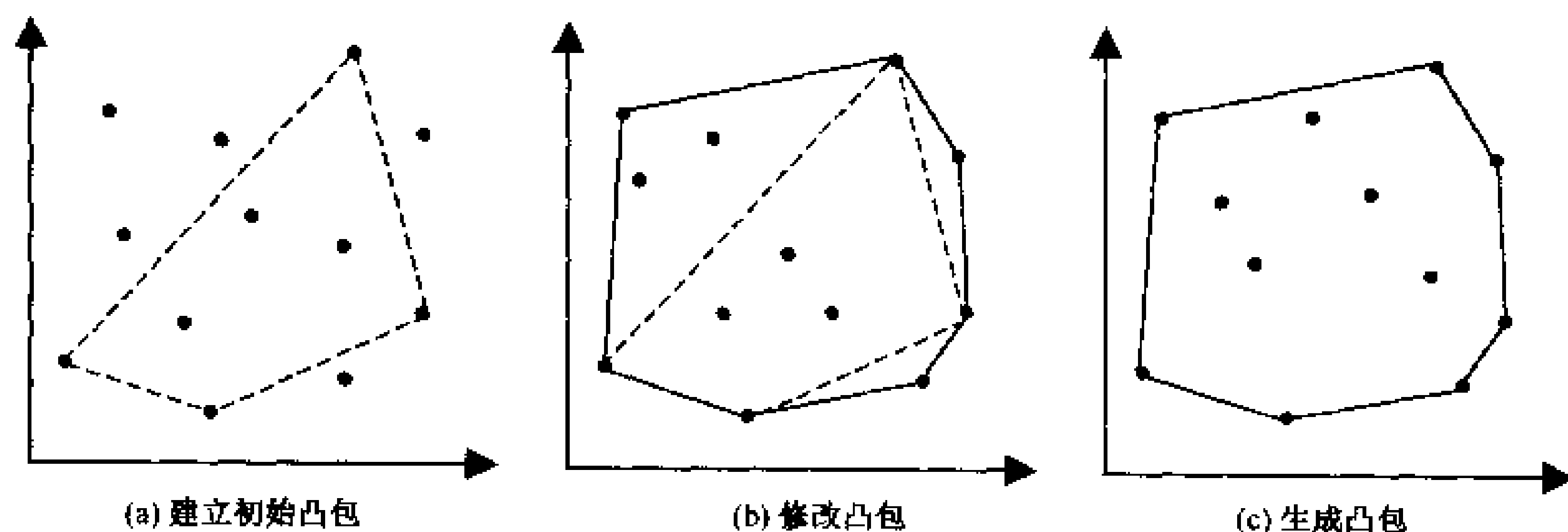


图 15-6　凸包的建立过程

（1）建立初始凸包：将 min$x$、min$y$、max$x$、max$y$ 的点或 min $(x+y)$、min $(x-y)$、max $(x+y)$、max $(x-y)$ 的点连接起来作为初始凸包，并按逆时针方向存储于链表中；

（2）修改凸包：对链表中相邻的两点，检查位于其右侧距离最大的点（计算三角形面积即可）。若点存在，或最大距离为零且该点位于上述两点之间，则将该点插入链表中上述两点之间；否则，不插入。重复上述过程，直到链表中的每一条边均判断一次，即完成凸包的形成；

（3）生成凸包：以最终形成的凸包代替初始凸包，并将链表写入凸包坐标数据文件。

基于凸包构建 D-TIN 有多种方法，典型的为前沿推进法和环切边界法。前沿推进法是以凸包的每一条边作为新生成三角形的起始边（所有三角形的边按逆时针方向存储），向凸包内逐渐“推进”，找到相应一点满足 Delaunay 构网法则，生成第一层 Delaunay 三角形。之后，依次以第一层三角形的各边为起始边，找到相应一点满足 Delaunay 构网法则，生成第二层 Delaunay 三角形……如此进行下去，直到构网结束（每条新边最多参与两次构网）。环切边界法是在凸包链表中每次寻找一个由相邻两条凸包边组成的三角形，在该三角形的内部和边界上都不包含凸包上任何其他点；将这两条凸包边的公共点去掉，得到新的凸包链表。重复以上过程，直到凸包链表中只剩下 3 个离散点为止。将凸包链表中的最后 3 个点构成一个三角形，该三角形连同已经找到的三角形形成整体，成为凸包三角形剖分结果。

基于凸包构建 D-TIN 之后，还可以在其中插入任意的特征点。即通过 Delaunay 法则、外接影响圆分析和分块结构加速搜索，生成加入点的影响 TIN（influence triangulation），来实现对 D-TIN 的局部更新和相邻拓扑关系的重新定义。也可采用逐点插入法进行 D-TIN 的局部重构。以图 15-7 为例，其插入重构的基本过程为：

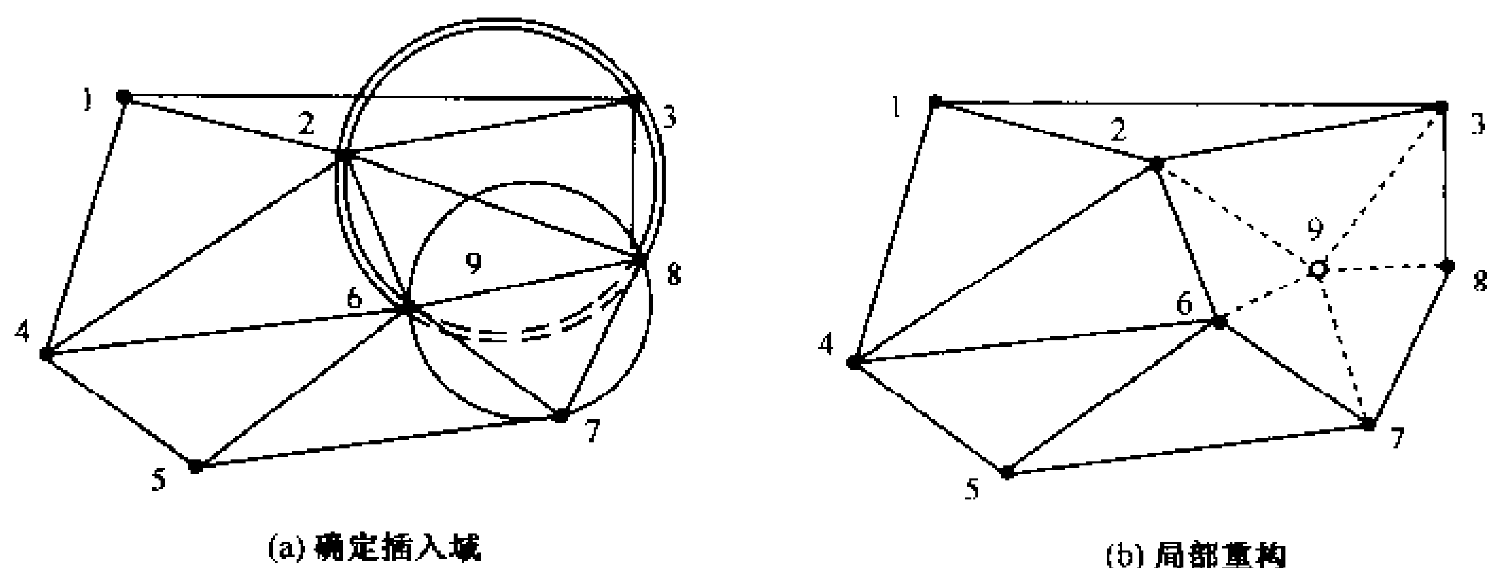


图 15-7　D-TIN 局部重构实例

(1) 找出外接圆包含插入点的所有三角形，构成插入区域；

(2) 删除插入区域内的三角形公共边（如图 15-7（a）中所示虚线），形成由影响三角形的顶点构成的多边形；

(3) 将插入点与该多边形的所有顶点相连，构成新的 Delaunay 三角形(如图15-7(b)中所示虚线)；

(4) 重复第（1）～（3）步，直到所有非凸包点都插入完。

15.2.2　有约束构 TIN 算法

在许多实际情况中，当利用离散点构建 TIN 时，不仅对三角形的形状有要求，而且对离散数据本身也有特殊要求。比如，某些点的连线（跨河大桥、地理边界、断裂线、结构线、河流等）对 TIN 网的局部合理性有决定性影响，某些点（山脊点、山谷点、断层点、河岸、湖岸等）必须连成一条线或一条封闭曲线等。此时，需要考虑将这些离散点给以某种强制约束，使得构 TIN 后符合实际情况，并提高 TIN 的质量。在二

维情况下，这种特征约束可以归纳为两种情况：一种是离散数据带有若干组“有向折线”，另一种是离散数据带有若干组“封闭多边形环”。通常，将上述问题称为基于矢量的有约束条件的离散数据构 TIN。

基于矢量的有约束的 TIN 构建方法有多种，代表性的有“两步法”。两步法的实质是：首先对约束数据集建立非约束 D-TIN（初始三角网），然后引入约束线段以嵌入之（调整过程）。根据引入约束条件的过程与方法的差异，“两步法”有许多不同的具体算法：① Bernal 和 Sloan 的算法在初始三角网过程中不区分约束与非约束点，而在调整过程中使之满足约束条件；利用连续的对角线交换法实现约束线段的嵌入，要求可交换对角线的四边形是严格凸的，而这一点在实际应用中较难满足；②'Floriani 算法的核心是简单多边形的 Delaunay 三角剖分，实现过程是递归执行，因而更具灵活性。其初始三角剖分网中可包含有带约束关系的散点，也可不包含。如果在初始三角剖分网中含有带约束关系的散点，则在约束线段的嵌入过程中，按散点的点号进行调整；如果不包含，则首先插入散点，然后调整约束关系。

以 Bernal 和 Sloan 的“对角线交换循环算法”为例，其基本原理是：首先，不考虑约束条件而构建初始 D-TIN；然后检测约束边所经过的所有三角形（其集合称为该约束边的影响域，影响域的每一条三角形边称为对角线）；从约束边的起始点出发，按一定的规则逐步交换对角线，最终使起始点和目标点相连。

此算法的关键是从起始点出发，要对遇到的每条对角线做可交换性判断。可交换就交换，不可交换就判断下一条。第一轮交换结束后，从头再来，开始下一轮。直到所有约束边均作为三角形边加入到 TIN 网中。

此算法的过程如下：

(1) 不考虑约束条件，构建初始 D-TIN；

(2) 检测并记录每条约束边的影响区域及其区域内的对角线；

(3) 从某一条约束边的起点 $firstP_i$（$i=1\sim n$，n 为约束边总数）开始，查询与该约束边相交的边 e_{ij}（$j=1\sim m$，m 为该边影响区域内的对角线总数），检测该边的相邻三角形的对点 $opptP_{i1}$是否为目标点 $lastP_i$：若是，则转入第 7 步；若非，则记录原始起点 $ofirstP_i = firstP_i$；

(4) 分别计算 e_{ij}的两个端点 Ne_{ijb}、Ne_{ije}与起始点 $firstP_i$、对点 $opptP_{i1}$所构成的三角形（$firstP_i$-Ne_{ijb}-$opptP_{ij}$，$firstP_i$-Ne_{ije}-$opptP_{ij}$）的面积（上述三点按逆时针方向排列时面积为正；反之为负），分别记录为 flag1 和 flag2。若一正一负，则转第 7 步；若都为正，则转第 5 步；若都为负，则转第 6 步；

(5) 将 Ne_{ijb}的值赋给 $firstP_i$，并检测此时临时起点 $firstP_i$ 与目标点 $lastP_i$ 记录是否为直接 TIN 边：若非，joinFlag 为真，转第 3 步；若是，则 joinFlag 为伪，则令 $firstP_i = ofirstP_i$，转第 3 步；

(6) 将 Ne_{ije}的值赋给 $firstP_i$，并检测此时临时起点 $firstP_i$ 与目标点 $lastP_i$ 记录是否为直接 TIN 边：若非，joinFlag 为真，转第 3 步；若是，则 joinFlag 为伪，则令 $firstP_i - ofirstP_i$，转第 3 步；

(7) 对角线交换，修改相应的拓扑结构；若 joinFlag 为真，记录 joinFlag 为伪，并将 $ofirstP_i$ 赋给 $firstP_i$，转第 3 步。

(8) 结束。

图 15-8、图 15-9 所示为两个具体例子。所不同的是，前者没有出现临时起点 first P_i 与目标点 lastP_i 记录为直接 TIN 边的情况，如图 15-8（c）所示；而后者则出现了临时起点 firstP_i 与目标点 lastP_i 记录为直接 TIN 边的情况，如图 15-9（c）所示。

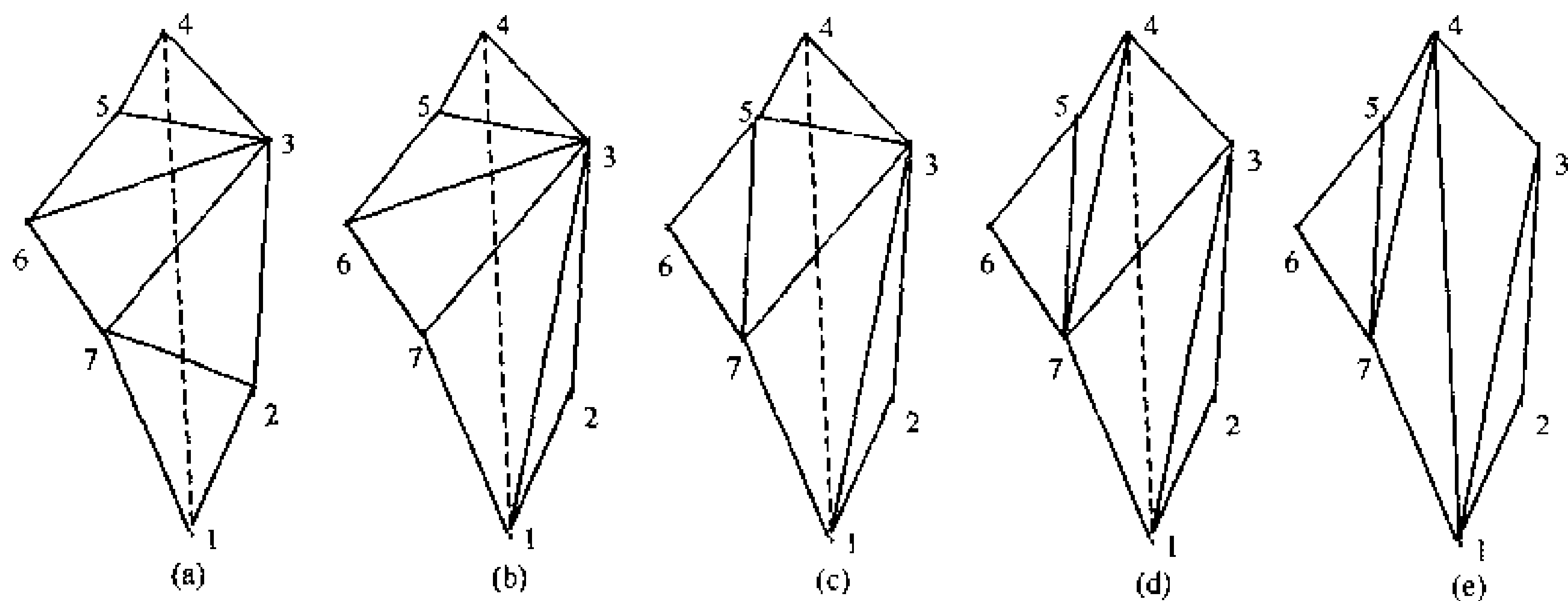


图 15-8 "多对角线交换循环算法"示例

(a) 开始；(b) 边 27 的两个端点 2、7 与起始点 1、对点 3 所构成的△123、△173 的面积分别为正和负，则交换 27 为 13；(c) 边 73 的两个端点 7、3 与起始点 1、对点 6 所构成的△176、△136 的面积均为正，则令 7 为起点；边 36 的两个端点 3、6 与起始点 7、对点 5 所构成的△735、△765 的面积分别为正和负，则交换 36 为 57；(d) 边 35 的两个端点 3、5 与起始点 1、对点 4 所构成的△134、△154 的面积分别为正和负，则交换 35 为 74；(e) 边 37 的两个端点 3、7 与起始点 1、对点 4 所构成的△134、△174 的面积分别为正和负，则交换 73 为 14，结束

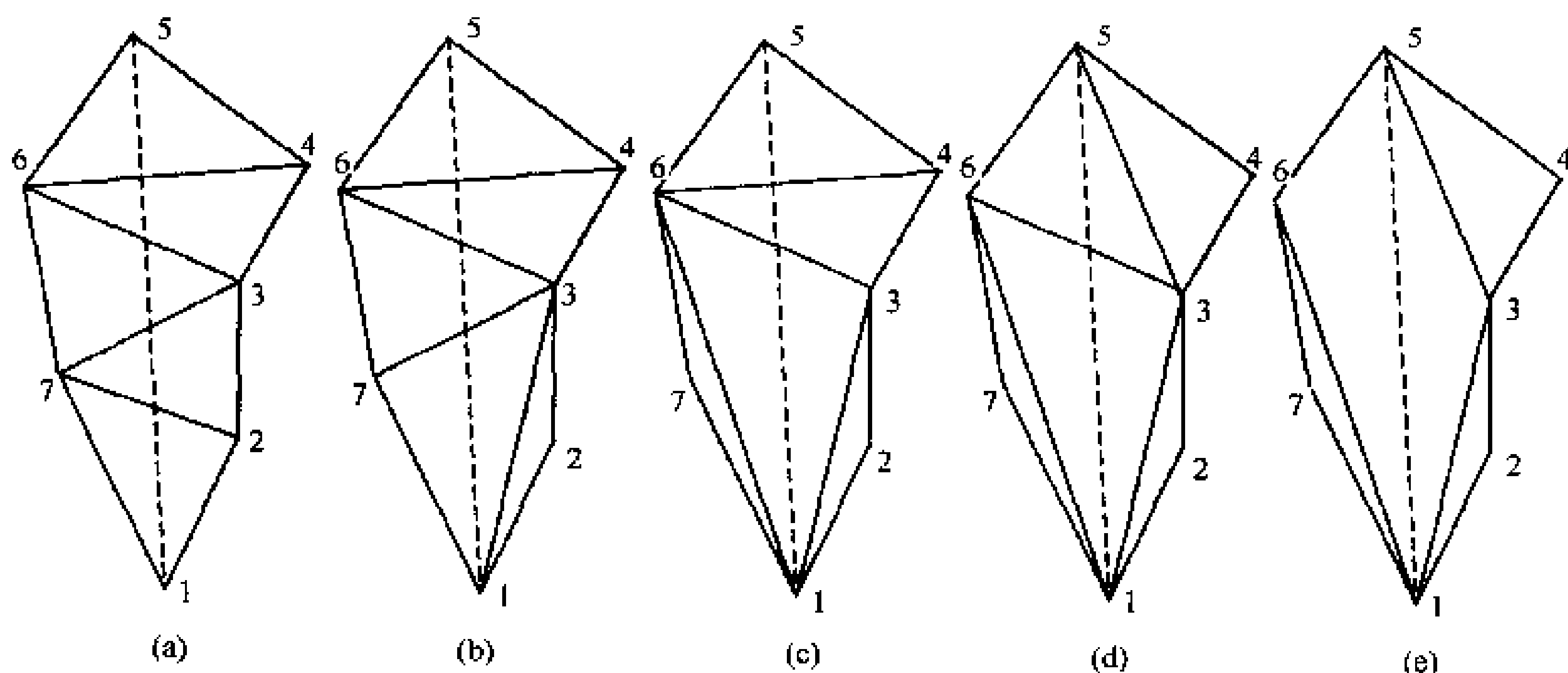


图 15-9 "多对角线交换循环算法"示例

(a) 开始；(b) 边 27 的两个端点 2、7 与起始点 1、对点 3 所构成的△123、△173 的面积分别为正和负，则交换 27 为 13；(c) 边 73 的两个端点 7、3 与起始点 1、对点 6 所构成的△176、△136 的面积分别为正和负，则交换 73 为 16；(d) 边 36 的两个端点 3、6 与起始点 1、对点 4 所构成的△134、△164 的面积均为负，则令 6 为起点；此时，6 已可与终点 5 直接连线为初始 TIN 的一条边，故 6 无效，回到起点 1 重新开始；边 64 的两个端点 6、4 与起始点 1、对点 5 所构成的△165、△145 的面积分别为正和负，则交换 64 为 53；(e) 边 63 的两个端点 6、3 与起始点 1、对点 5 所构成的△165、△135 的面积分别为正和负，则交换 63 为 15，结束

由此可见，基于矢量方式的约束 TIN 构建的大量计算是判断初始 Delaunay 三角形的边是否相交。当约束线较长（其上的约束点较多），而且整个格网的数据量较大时，此过程将很花时间。因此，有人提出另一种方法，即直接考虑约束线上的点来生成 D-TIN，而不是通过对角线交换来不断修正。

§15.3 基于等高线的构 TIN 算法

15.3.1 基于等高线构 TIN 的一般算法

等高线是对离散采样点经内插、光滑等处理后得到的一种表达地面高程与地形特征的抽象方式，是一种传统的地图制图方法。由于等高线难以直观地表达地形的起伏变化与地形特征，非专业的人员在阅读和利用等高线图时往往存在一定困难。随着计算机应用技术和 GIS 技术的发展，人们已乐于使用 TIN 或格网模式来模拟和表达地面高程特征。前已述及，采用离散采样点来构建 TIN 是一种直接的方式，但在很多情况下，已经不再保留原始采样数据，而只有等高线图或数字化的等高线数据。如何利用等高线数据来构建 TIN，尤其是构建高质量的 TIN，就成为 GIS 中的一项重要任务。

从等高线生成 TIN 一般有三种方法，即等高线的离散点直接生成法、加入特征点的 TIN 优化法和以等高线为特征约束的特征线法。

1. 等高线的离散点直接生成法

该法是直接将等高线上的点离散化，即按一定间距从等高线上重采样（因此也可称为等高线重采样法），或从 CAD 等软件系统所管理的等高线文件中读取离散点数据，然后基于重采样点（或离散点）构建 D-TIN。由于仅考虑了点而忽略了等高线数据的特殊几何结构，往往会导致很坏的结果，如在谷区、峰区或大拐弯处出现三角形的三个顶点均位于同一条等高线上（即所谓的“平三角形”），或者三角形的某一条边穿越了等高线等。而以上情况都是不允许的，因此，这种直接法在实际工作中很少采用。

2. 加入特征点的 TIN 优化法

这种方法实质上是对直接生成法的改进。在等高线离散点构 TIN 的基础上，采用加入特征点的方式来消除“平三角形”，并按某种准则通过扫描优化方式来消除穿越等高线等不合理现象。特征点的加入可以是手工加入，但在效率、合理性、完整性等方面存在问题；也可以设计一定的算法，由计算机来自动提取特征点，算法的原理主要是基于原始等高线的拓扑关系。

3. 以等高线为特征约束的特征线法

该法的核心思想是：每一条等高线必须当作特征线或结构线，而且线上不能有三角形生成，即三角形不能跨越等高线。基于这一思想，许多研究者提出不同实现算法，如 Gannapathy 和 Denneby（1982），Christensen（1987）等。

无论是基于等高线图，还是基于数字化的等高线数据，以等高线为特征约束的特征

线法均有一个数据预处理的过程。预处理的主要内容包括：数据数字化、离散化，离散数据点分布均匀化，地形特征点（即地面曲率变化点和坡度变化点，如峰点、谷点、鞍点、变坡点等）与特征线（山脊线，山谷线或流水线等）的加入，以及地形突变线（断层、陡坎、悬崖等）与突变区（陷落柱、岩溶柱、孤峰、洼地等）的加入等。

根据是否加入地形特征点与特征线、以及是否加入地形突变线与突变区等，可以将基于等高线构建 TIN 的算法分为有约束和无约束两种基本模式。显然，以等高线为特征约束的特征线法要求所构建的三角形不可跨越等高线，即等高线本身就是约束条件，因此，从本质上说以等高线为特征约束的特征线法属于约束条件下离散点的三角剖分。

15.3.2 基于矿层底板等高线的构 TIN 算法

矿层底板等高线图作为矿山企业日常生产必备的基础图纸之一，在矿井生产、储量计算和矿山总体规划等多方面起着非常重要的作用，而现有矿层底板等高线图仍然停留在纸图或 CAD 图的基础上。这些图纸无法满足矿山发展对数据共享、数据挖掘的要求，也无法直观的表达矿层底板的形态特征和各个矿层之间的相互关系，更无法适应“数字矿山”（Digital Mine, DM）的战略实施。因此，以现有的纸图或 CAD 图为基础，采用一种新的方式表达煤层底板的形态成为当务之急。

矿层底板等高线不同于一般地形图的等高线，有其特殊性。首先，矿层底板等高线是矿层底板标高相同的点在零水平面上的投影图，由于矿层赋存区域的不规则，底板等高线图亦呈不规则性，而且多不封闭，无规律可循；其次，矿层底板等高线的绘制方法不同于其他等高线图的绘制，只能依靠有限的资料来描绘等高线图，常规方法是利用钻孔数据采用剖面法绘制，因而其数据来源十分有限。此外，矿层底板等高线是与断层数据紧密结合的，断层数据作为一种强约束条件出现。

无论是利用底板等高线纸图还是已有底板等高线图的 CAD 文件为数据源，均需实现平面二维数据到三维数据的转化。可利用 AutoLisp 语言编程获取等高线 CAD 文件数据。矿层底板等高线数据和断层数据的特点是约束条件复杂、离散点少，故采用 Sloan 两步法进行矿层底板 TIN 模型的构建，并利用改进的连续的对角线交换法来实现任意条件下约束条件的嵌入（郝海森等，2002，2003）。

由于断层的形状和空间分布很复杂，既有无推掩关系的正断层，也有有推掩关系的逆断层，还存在既有推掩关系又有走滑关系的复杂断层。如何使用 TIN 来反映其真实形状，是问题的关键。

1. 构模流程

根据 CD-TIN 分步法构网特点，底板 TIN 模型构建的技术流程框图如图 15-10 所示。

在构 TIN 过程中需建立点、线的索引以及三角形的编号和拓扑关系。在此过程中主要有两个查询操作：一是查询与边构成三角形的第 3 点；二是查询新生成的边在已生成的边中是否存在。所以，建立点、线的索引对于构建 TIN 的构网速度有很大影响。现有建立索引的方法主要包括两种：一种是将所有离散点按 x 和 y 的大小排序，利用

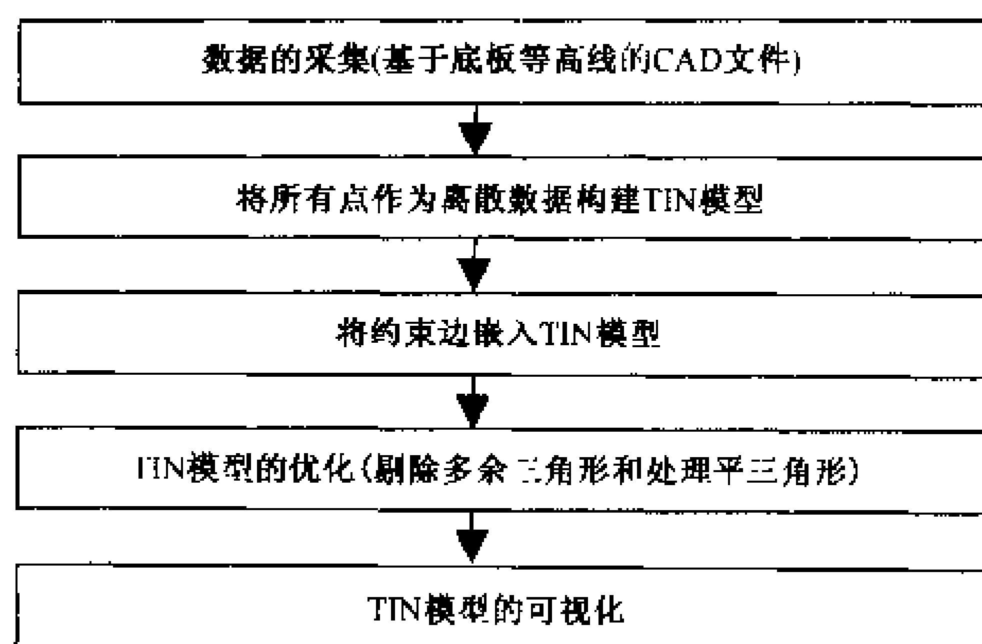


图 15-10 矿层底板 TIN 模型构建流程

折半查找的方法快速查找构网的第 3 点；二是按分区建立索引，它是将所有离散点分成若干个区域，每个区域包括若干点。只需知道所查点的坐标，算出此点的分区数，即可快速获得其相邻网格内的所有点。因此，建立查询速度快、效率高的索引机制就成为此算法的关键。我们采用分区索引，实现了高效查询。

2. 断层处理

一般来说，断层分为单一正断层、单一逆断层、正断层和正断层切割、正断层和逆断层切割、逆断层和逆断层切割几种典型形状，我们利用矿层底板等高线图的断层信息（包括断层的属性、上下盘信息、断层上点的坐标和高程等），采用 TIN 模型，来描述断层的真实形态。

1）单一正断层

对于单一正断层，其上翼（盘）相对于下翼（盘）沿断裂面向下移动。在等高线图上两者的位置并没有发生变化，将断层数据嵌入后所形成的三角网可以完全表示其模型，故不需进行断层影响区域的局部构 TIN 和其他处理。

2）单一逆断层

对于单一逆断层，因为逆断层的上翼沿断裂面相对于下翼向上移动，上盘上升，上下盘相对错动。这样，在等高线的水平投影图中，上下盘位置与实际位置相比发生交换，将断层数据嵌入 TIN 模型后所形成的三角网并不能完全表示其形态。为了解决这一问题，需在断层的影响区域内，分别以上下盘数据和其在空间实际相连接的点数据重新分区，然后局部构 TIN。如图 15-11 所示。

3）正断层和正断层相切割

正断层和正断层相切割，两断层的上盘分别相对其下盘沿断裂面向下移动，在等高线图上的投影没有发生位置的交换。这样，将断层约束加入 TIN 模型后，即可表示两

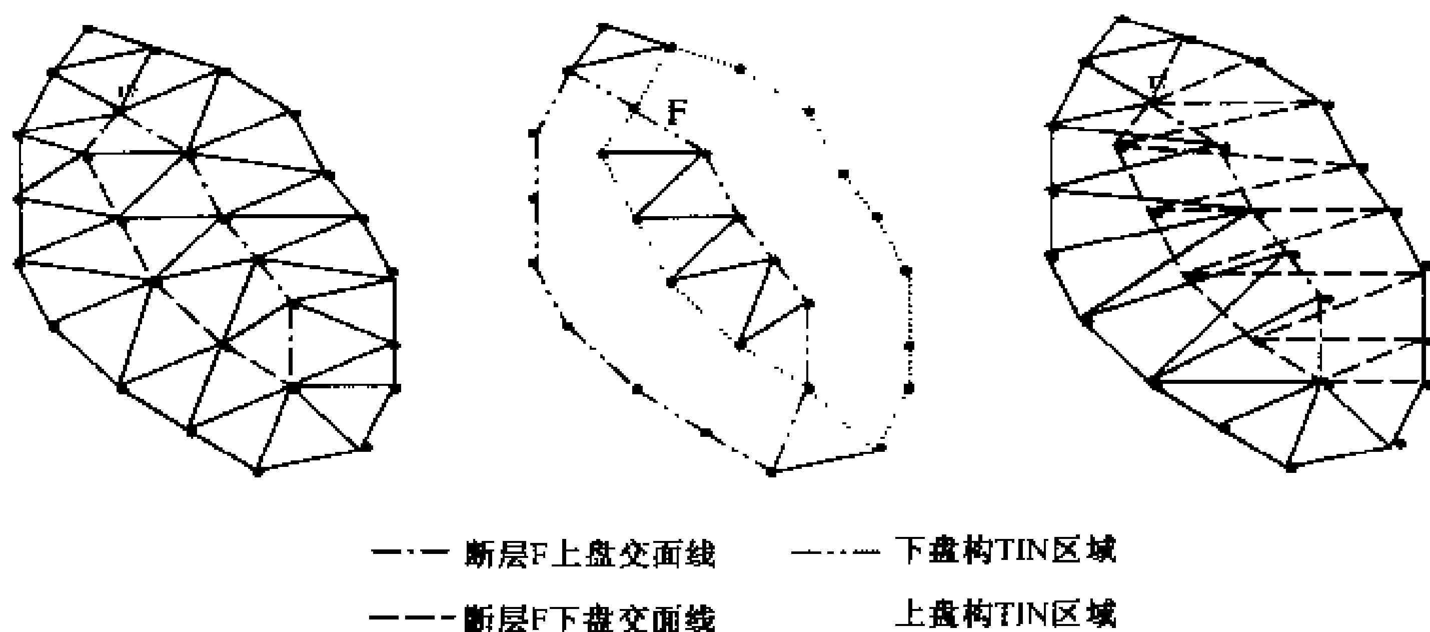


图 15-11 单一逆断层嵌入 TIN 模型及分区构 TIN
（郝海森等，2002）

正断层相割的真实状态，无须局部构 TIN。

4）正断层和逆断层相切割

正断层和逆断层相割时，正断层的上盘相对其下盘沿断裂面向下移动，在等高线图的投影上没有发生位置的交换。而逆断层的上盘相对其下盘沿断裂面向上移动，这样导致逆断层在等高线上的投影位置发生交换。将断层数据嵌入 TIN 模型后，仍需对断层所在影响区域作一定的分区处理，然后局部重新构建 TIN 模型。如图 15-12 所示。F1 是逆断层，F2 是正断层。左图为分区示意图，依据两断层特点将影响区域分成四部分，分别是 *e*-1-*c*-*e*、*f*-2-*a*-*f*、*b*-3-*h*-*b*、*g*-4-*d*-*g*。这四个区域内分别局部构 TIN。右图是 *e*-1-*c*-*e* 和 *b*-3-*h*-*b* 两局部构 TIN 示意图。

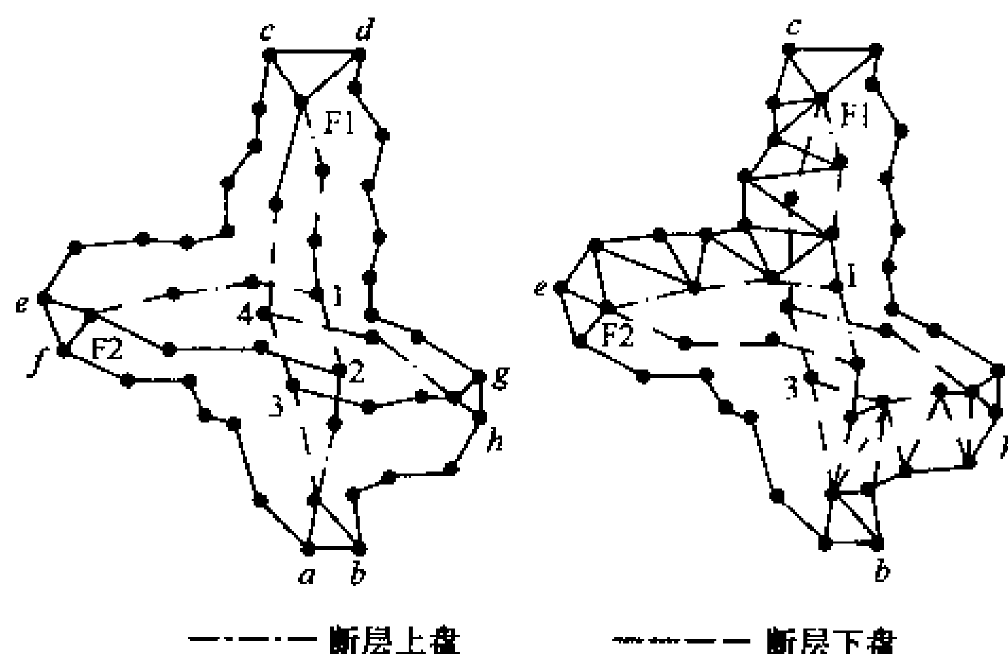


图 15-12 相切割的正断层和逆断层的局部构 TIN
（郝海森等，2002）

5）逆断层和逆断层相割

逆断层和逆断层相割的情况较为复杂。两个断层的上盘分别相对其下盘沿断裂面向

上运动。这使得两断层的上下盘在等高线图上的投影位置都发生交错，即使将断层数据作为约束嵌入三角形网中，也不能正确表达断层的实际形态。为此，依据逆断层的特点，在两断层的影响区域内采用局部构网的方法实现断层 TIN 模型的重构，如图 15-13 所示。左图是分区示意图，同样分为 4 部分，分别是 *c*-1-2-*g*-*c*、*h*-3-4-*a*-*h*、*b*-3-4-*f*-*b* 和 *e*-1-2-*d*-*e*。右图是 *e*-1-2-*d*-*e* 和 *h*-3-4-*a*-*h* 两区域的局部构 TIN 的示意图。

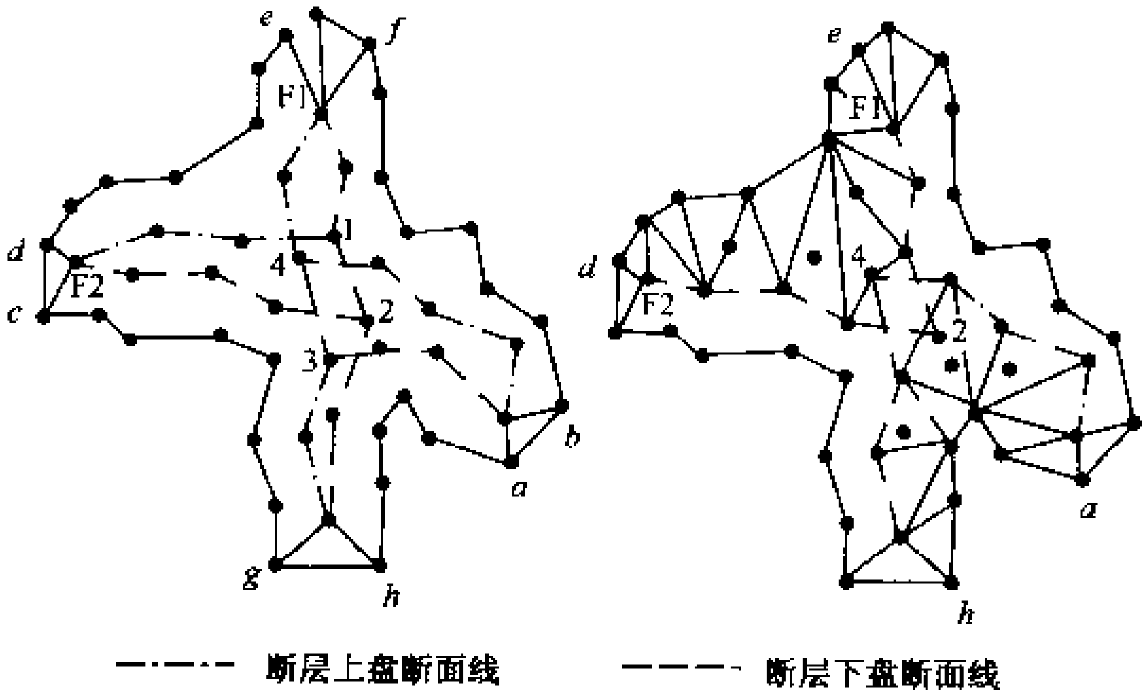


图 15-13　相切割的两逆断层的局部构 TIN
（郝海森等，2002）

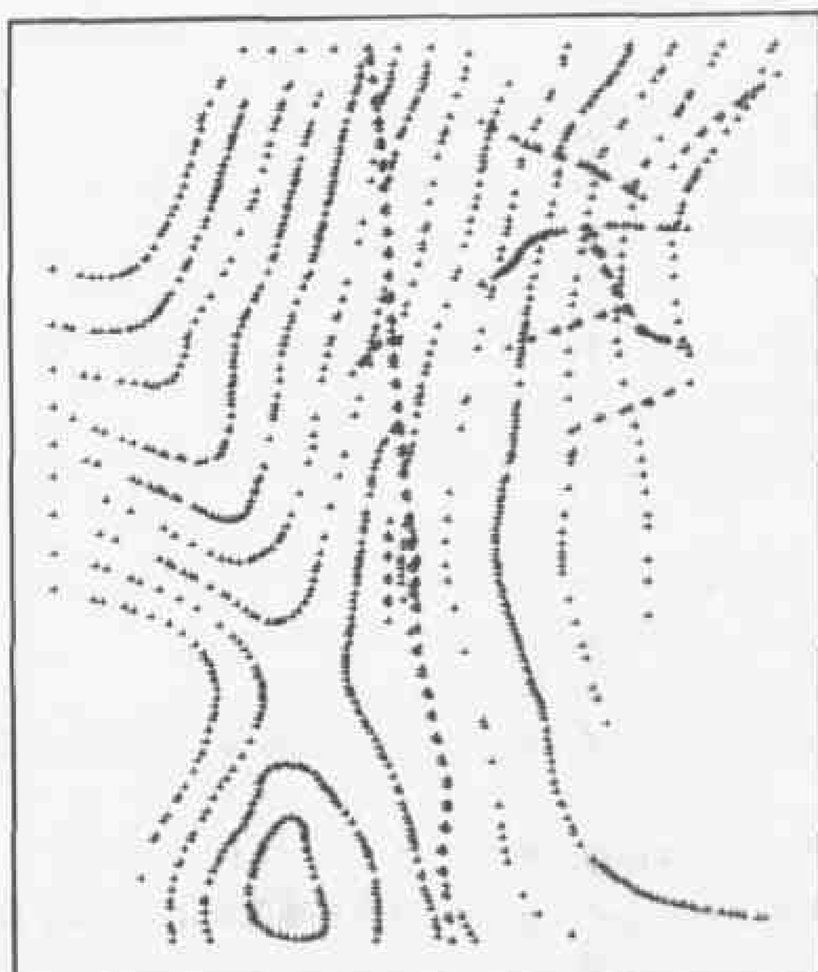
通过以上分析可以看出，利用断层数据嵌入 TIN 模型后所形成的影响区域，采用分区局部重构 TIN 的方法可以实现基于底板等高线矢量数据对典型断层的正确描述。

3. 程序实现和煤层可视化

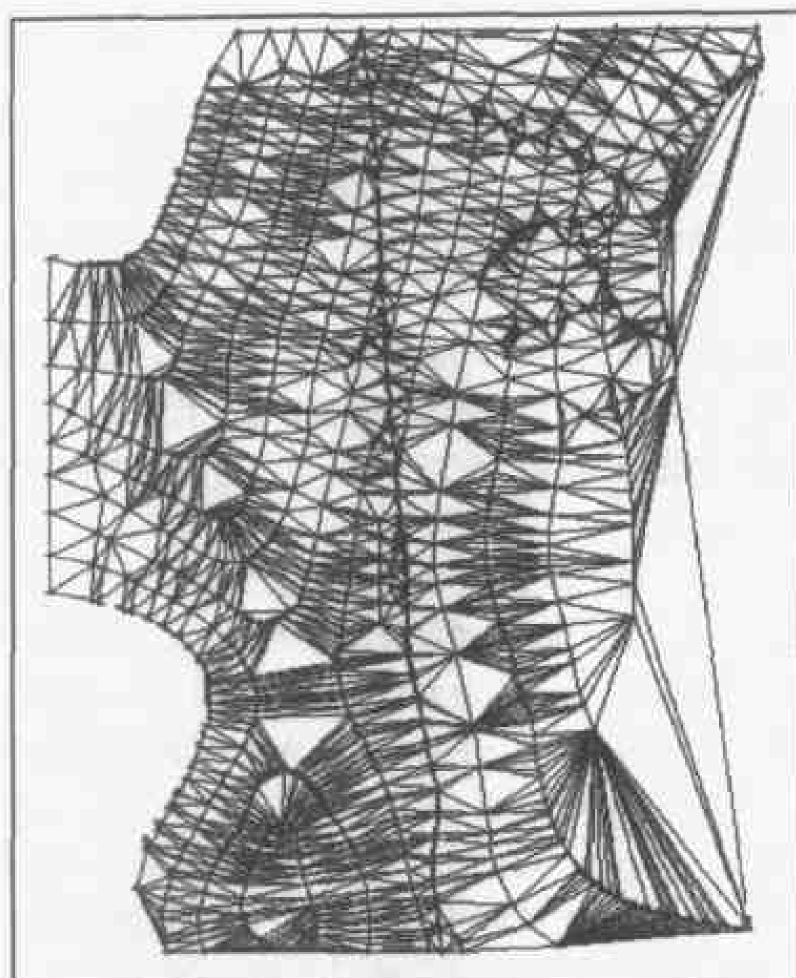
基于 Delhi 和 OPGL 开发工具，在所开发的三维地学模拟系统——GeoMo^{3D}中开发了插件。以河北某煤矿多断层发育的煤层底板等高线 CAD 文件作为实验数据，效果较好。此程序能够很好地解决矿层底板的数字化，建立以三角网为基础的矿层底板表面模型，实现包括逆断层等多种典型断层的模型表达，并解决了脊线平台化、TIN 丛集化等问题。脊线平台化问题是通过沿脊线插入离散点来局部重构 TIN，插入点的方式分为计算机自动插点和人工插点；丛集化问题的出现是由于相邻等高线采样数据的不均匀，有些点相距过远，解决的办法是改良预处理策略。

图 15-14 中的两图分别是等高线离散点提取结果和强约束 D-TIN 建模结果，其中包括一条主断层和多条小断层。

图 15-15（a）是单煤层 DEM 可视化的效果。由于每个钻孔都对应一个 Voronoi 区域、地层厚度、矿层容重等参数，因此可以在按钻孔生成的 Voronoi 图的基础上，任选一多边形区域，依据此多边形区域和 Voronoi 图的交集，还可计算该区域内的地层体积或矿层储量，如图 15-15（a）中 Voronoi 图所示。图 15-15（b）是在图 15-15（a）的基础上，进行该区域多煤层 DEM 可视化的效果。



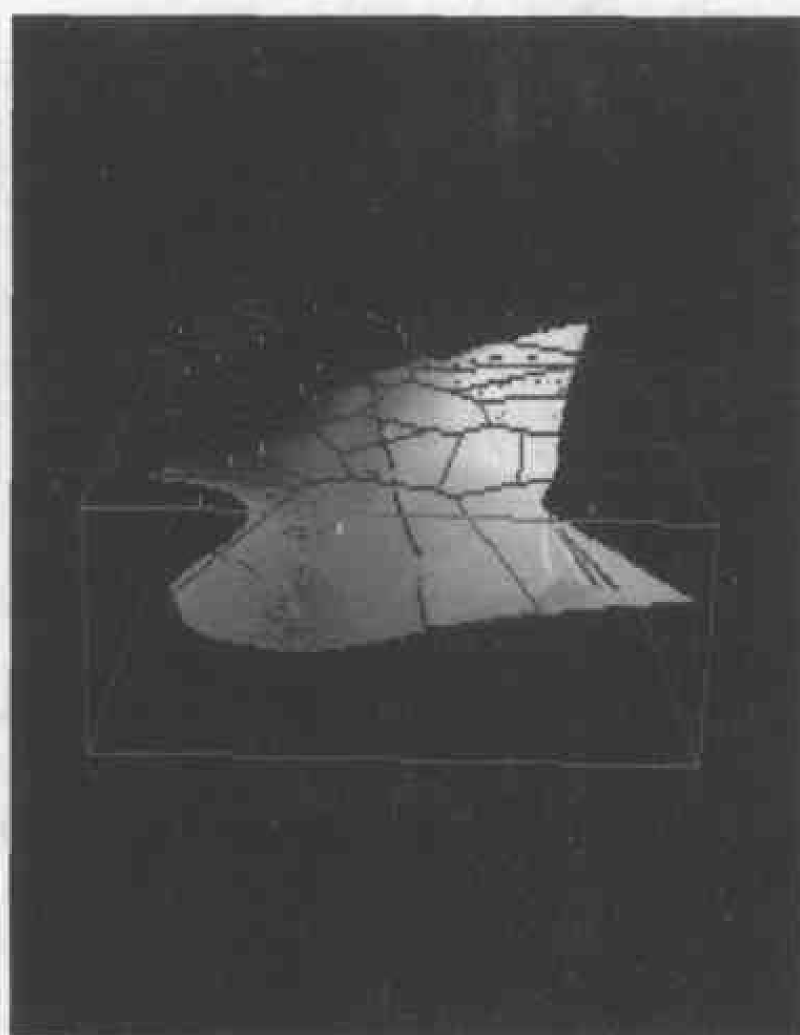
(a) 基于底板等高线的离散点重采样结果



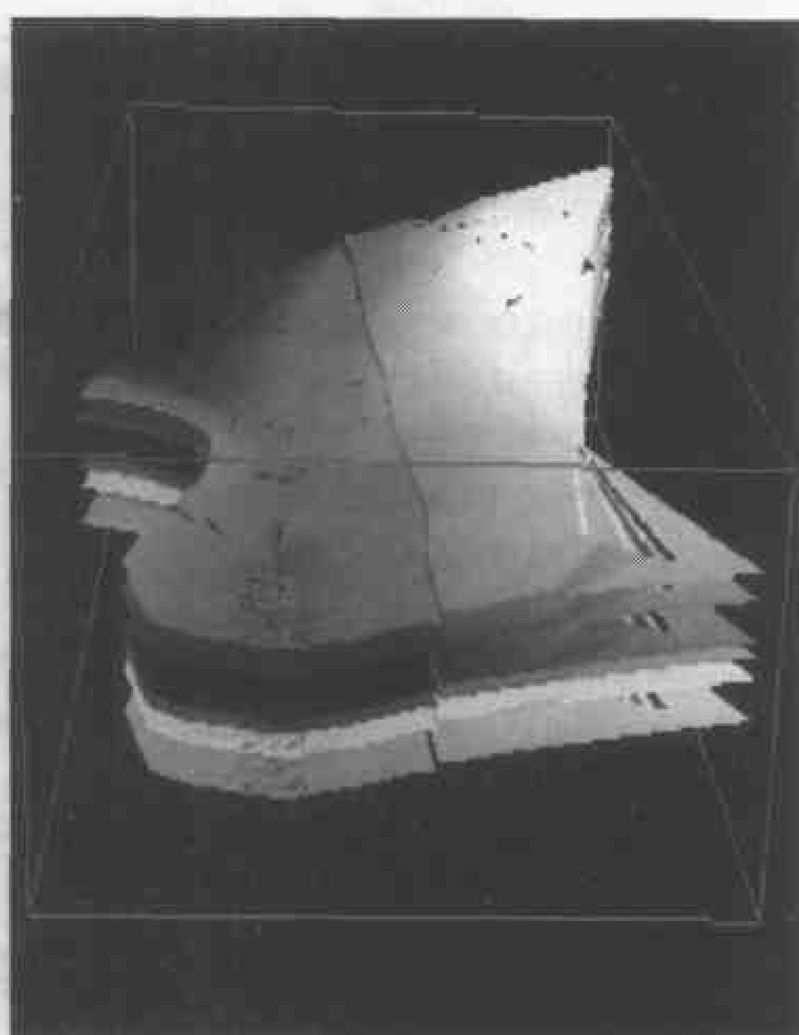
(b) 煤层底板强CD-TIN

图 15-14　河北某矿的矿层底板等高线构 TIN 实例

（郝海森等，2002）



(a) 单煤层可视化及储量计算



(b) 多煤层组合DEM效果

图 15-15　煤层底板 DEM 可视化结果

（郝海森等，2003）

参 考 文 献

邸 元. 2000. 基于等高线建立DTM中平坦区域的一种处理方法. 计算机辅助设计与图形学报, 12 (8): 566～570
郝海森, 吴立新. 2002. 构建含断层的煤层底板约束TIN模型, 第六届全国矿山测量学术会议论文集: 32～35
郝海森, 吴立新. 2003. 基于强约束Delaunay-TIN的三维地学模拟与可视化. 地理与地理信息科学, 19 (2): 15～18
姜寿山等. 1995. 用空间形状优化标准完成散乱数据的三角剖分. 计算机辅助设计与图形学学报, 7 (4):
柯映林. 1992. 离散数据的几何造型技术及其应用研究 [博士学位论文]. 南京航空航天大学
李立新, 谭建荣. 1999. 约束Delaunay三角剖分中强行嵌入约束边的多对角线交换方法. 计算机学报, 22 (10): 1115～1118
李志林, 朱 庆. 2000. 数字高程模型. 武汉: 武汉测绘科技大学出版社, 34～56
凌海滨, 吴 兵. 1999. 改进的自连接Delaunay三角网生成算法. 计算机应用, 19 (12): 10～12
刘学军, 龚健雅. 2001. 约束数据域的Delaunay三角剖分与修改算法. 测绘学报, 30 (1): 82～83
卢朝阳, 吴成柯. 1997. 任意多边形内带特征约束的散列数据的最优化三角剖分. 计算机辅助设计与图形学报, 9 (4): 302～308
马小虎等. 1997. 一种基于图的平面点集Delaunay三角剖分算法. 中国图像图形学报, 2 (1): 7～11
马小虎, 潘志庚, 石教英. 1999. 基于凸凹顶点判定的简单多边形Delaunay三角剖分. 计算机辅助设计与图形学报, 11 (1): 1～3
毛善军, 马宏兵. 1999. 自动构建复杂地质体数字高程模型的方法研究. 测绘学报, 28 (1): 57～61
闵卫东, 唐泽胜. 1995 二维任意域内点集的Delaunay三角划分的研究. 计算机学报, 18 (5): 357～364
宋占峰. 2000. 不规则三角网的数模的快速搜索与定位, 18 (2): 31～34
王家耀. 2001. 空间信息系统原理. 北京: 科学出版社, 172～186
王永明. 2000. 地形可视化. 中国图象图形学报, 5 (6): 449～456
吴立新, 殷作如, 邓智毅等. 2000. 论21世纪的矿山——数字矿山. 煤炭学报, 25 (4): 337～342
武晓波, 王世新, 肖春生. 1999. Delaunay三角网的生成算法研究. 测绘学报, 28 (1): 28～35
徐春蕾, 李思昆. 2000. 一种适用任意平面多边形的三角剖分算法. 国防科技大学学报, 22 (2): 82～85
杨晓东, 刘春太, 申长雨等. 2000. 内部带特征约束的任意平面域的三角形网格生成方法. 计算物理, 17 (3): 293～297
周立吾, 张国良, 林家聪. 1987. 矿山测量学. 徐州: 中国矿业大学出版社
朱晓岚, 何新义, 陈于恒. 1987. 矿体几何学. 徐州: 中国矿业学院出版社
朱 庆, 陈楚江. 1998. 不规则三角网的快速建立及动态更新. 武汉测绘科技大学学报, 23 (3): 204～207
Anglada M. V., 1997. Improved incremental algorithm for constructing restricted Delaunay triangulations. Computers & Graphics, 21 (2): 215～223
Christoper G., Dakowicz M., 2001. Terrain Modelling From Contours. The 3rd ISPES Workshop on Dynamic and Multi-Dimendional GIS & The 10th Annual Conference of CPGIS on Geoinformatics: 427～432
Cignoni P., Montanil C. and Scopigno R. D., 1998. A fast divide and conquer Delaunay Triangulation algorithm. Computer Aided Design, 30 (5): 333～341
Floriani. L. D., Puppo E., 1992. An On-Line Algorithm For Constiained Delaunay Triangulation. CAGIP: Graphical Models and Image Processing, 54 (3): 290～300
Lawson C. L., 1977. A software for C surface interpolation. In: Rice J R. ed., Mathematical Software III. New York, Academic press: 161～94
Lee D. T., Schachter B. J., 1980. Two Algorithms For Constructing a Delaunay Triangulation. Int J of Computer and Information Science, 9 (3): 219～242
Sapidis N., Perucchio R., 1991. Delaunay triangulation of arbitrarily shaped planar domains. CAGD. 8 (6): 421～437
Watson D. F., 1981. Computing the n-dimentional Delaunay tessellation with applications to Voronoi polytopes. The Computer J. 24 (2): 162～166

第十六章　Voronoi 图构建算法

Voronoi 图又称为 Dirichlet 镶嵌（tessellation），在二维空间中也称 Thiessen 多边形。其概念由 Dirichlet 于 1850 年首先提出；1907 年，俄国数学家 Voronoi 对此作了进一步阐述，并提出高次方程化简。1911 年，气象学家 Thiessen 为提高大面积气象预报的准确度，应用 Voronoi 图对气象观测站进行有效区域划分。在计算几何上，Voronoi 图也称为 Delaunay 三角划分的对偶图。

1975 年，Voronoi 图被作为一种有效的数据结构引入计算机领域（Shamos M. & Hoey D.，1975），成为计算几何领域的主要研究热点之一。根据 Voronoi 图发生源的不同，可以分为平面点集 Voronoi 图、平面曲线与多边形 Voronoi 图、球面点线面集 Voronoi 图和空间多面体 Voronoi 图 4 种主要形式。这些不同形式的 Voronoi 图分别将平面、球面和空间根据距离的远近关系进行了一定的区域划分，每个分区分别从属于一个发生元素，并且分区内的点到该发生元素的距离比到其他元素近。

由于 Voronio 图无歧义地描述了空间对象的相对邻近关系，并具有动态性和层次性特点及其他奇妙的集合特性，在气象学、物理学、地质学、CAD/CAM、机器人运动规划、生态学、城市规划等领域得到了广泛应用。Voronoi 图已成为 GIS 基础理论与算法中的一项重要内容，在空间邻近关系建立、空间区域划分、空间邻近查询、空间邻近分析及专题制图等方面有重要的应用。比如，GIS 中空间目标之间的邻接关系可以在点、线、面之间的拓扑表中显式表达，而目标之间的邻近关系则要利用直线求交等技术来进行判断。若利用 Voronoi 图，就可以有效地进行目标的邻近关系分析。

§ 16.1　平面点集 Voronoi 图构建算法

平面点集 Voronoi 图也称常规 Voronoi 图（ordinary Voronoi diagram，OVD）、Thiessen 多边形或 Dirichlet 图。它是对平面的一个划分，每个分区表示一些点的轨迹，这些点到点集中的某个点比到其他点更近；它构建并表达相邻但不相连的平面离散点群实体之间的空间邻近拓扑关系。平面点集 Voronoi 图的构建方法有许多（Shamos M. & Hoey D.，1975；Hwang F.，1979），经典点集的算法归纳起来有 4 种，即插入法［图 16-1（a）］、间接法［图 16-1（b）］、分治法［Lee D. et al.，1981；图 16-1（c）］和扫描线法（Fortune S.，1986；Aurenhammer F. and Edelsbrunner H.，1984），时间复杂度最优达到 $O(N\lg N)$，存储空间最优达 $O(N)$。

平面点集 Voronoi 图的数学定义如下：

设研究区域有一组离散点（x_i，y_i）（$i=1, 2, 3, \cdots, k$，k 为离散点数），将该区域用一组直线段分成 k 个互相邻接的多边形，使得：

（1）每个多边形内包含且仅包含有一个离散点；

（2）位于包含离散点（x_i，y_i）的多边形内的任意一点（x，y），恒满足：

$$\sqrt{(x-x_i)^2+(y-y_i)^2} < \sqrt{(x-x_j)^2+(y-y_j)^2} \cdots\cdots (i \neq j) \quad (16\text{-}1)$$

(3) 若 (x, y) 位于离散点 (x_i, y_i) 和 (x_j, y_j) 所在多边形的公共边上，则以下等式恒成立：

$$\sqrt{(x-x_i)^2+(y-y_i)^2} = \sqrt{(x-x_j)^2+(y-y_j)^2} \cdots\cdots (i \neq j) \quad (16\text{-}2)$$

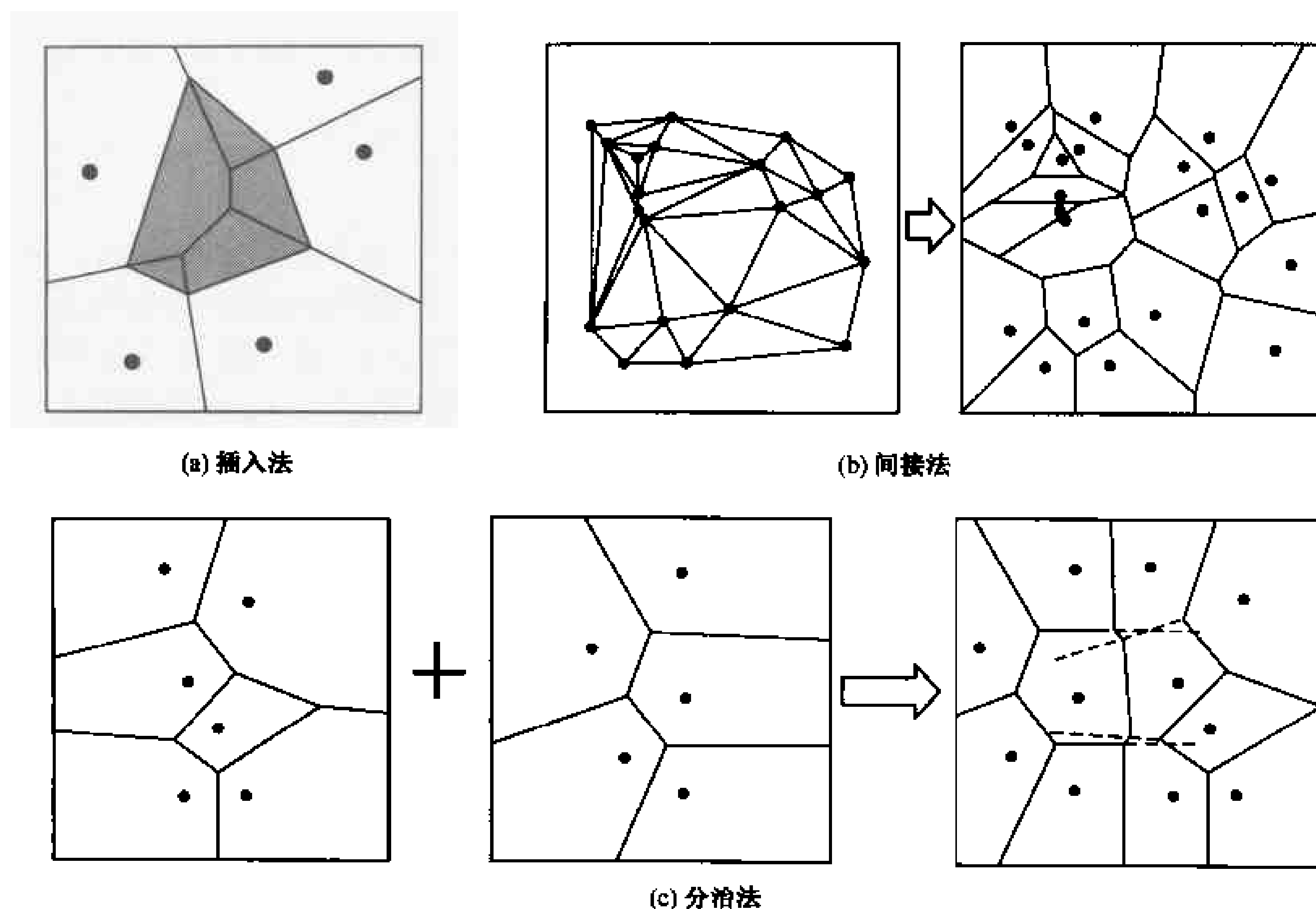


图 16-1 平面点集 Voronoi 图的生成算法分类
（据赵学胜，2002）

满足以上条件的多边形为该平面点集的 Voronoi 图。平面点集 Voronoi 图的基本性质有：惟一性；凸多边形；任意两个多边形不存在公共区域。

一般情况下，Voronoi 图的一个顶点同时属于 3 个 Voronoi 多边形或更多，每个 Voronoi 多边形内有且只有一个点。一个完整的 Voronoi 图应该由多个 Voronoi 多边形组成，其中第 i 个多边形的数学表达形式如下：

$$V_i = \{x \in R^2 : \| x - P_i \| \leqslant \| x - P_j \|, i = 1 \sim n, j = 1 \sim n, j \neq i\} \quad (16\text{-}3)$$

式中，$\| x - P_i \|$ 为平面域上点 x 与点元素 P_i 之间的欧式距离；$\| x - P_j \|$ 为平面域上点 x 与点元素 P_j 之间的欧式距离。

平面点集 Voronoi 图的构建可以直接基于离散点的矢量坐标来实现，也可以利用基于栅格的扩张原理来进行。

16.1.1 基于矢量的Voronoi图自动构建

1．平面点集 Voronoi 图的数据结构

按照平面点集 Voronoi 图的性质，为了满足 Voronoi 图的构建及邻近查询等要求，其数据结构可以设计为：

离散点数据结构：

```
struct ScatterPtObj {
    Pt3D;                                    //离散的坐标
    ID AdjacentPt;                           //关联的最邻近离散的标识号
    ID RelateTri;                            //关联的三角形标识号
}
```

平面点集 Voronoi 图数据结构：

```
struct VoronoiObj {
    ID Num;                                  //多边形标识号
    ID ScatterPt                             //多边形所包含的离散点
    Carray<Pt3D, Pt3D> VoronoiPolyPt;        //多边形顶
}
```

2．平面点集 Voronoi 图生成过程

给定一个 D-TIN，对于它的所有内边，连接共有每条内边的两个三角形的外接圆的圆心，即构成该 TIN 的平面点集 Voronoi 图。

(1) 首先构建离散平面点集的 D-TIN；

(2) 然后求取各三角形的外接圆心；

(3) 对每一个离散点，按顺时针或逆时针方向连接与其关联的三角形的外接圆心，即得到该离散点的 Voronoi 多边形；

(4) 将各离散点的 Voronoi 图形成集合，即得到本平面点集的 Voronoi 图。

所构建的平面点集 Voronoi 图是以平面点集凸包为外部边界的。通常，Voronoi 图的构建是在某一规则的有限边界（如图框、矩形）范围内，或在某一任意边界范围内（如地理边界、土地边界、行政边界等）进行的。因此，构建完平面点集 Voronoi 图之后，要以点集凸多边形边界所在的三角形的外接圆心分别向该凸多边形边界作垂直射线，射线与图框或任意边界相交，即得到凸包边界离散点的 Voronoi 多边形的有效区域，如图 16-2 所示。

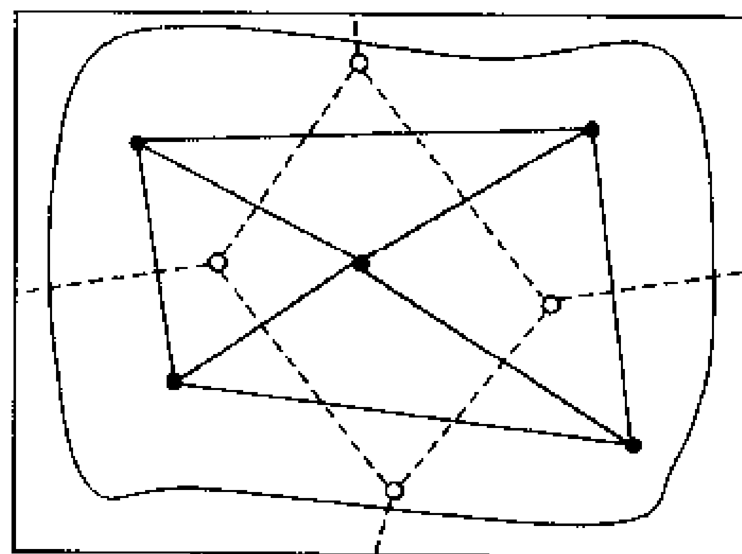


图 16-2 D-TIN、Voronoi 图与外部边界

3. 平面点集 Voronoi 图的任意多边形裁剪

实际上，研究区域的边界不仅可能是不规则的，而且可能不包含所有点集，即边界与点集凸包有相交之处。此时，有必要对业已形成的基于图框边界的点集 Voronoi 图进行边界裁剪，最后得到研究区域内的平面点集 Voronoi 图。

王家耀（2001）在 Sutherland 和 Hodgeman 的逐次裁剪算法的基础上，发展了一种适用任意裁剪窗口的算法。该算法的基本思想：以顺时针方向排列顶点序列的 Voronoi 多边形为输入，依次考虑 Voronoi 多边形的各条边的两个顶点，并判断这两个顶点与裁剪边界线的关系，裁剪后将窗口内的 Voronoi 图输出。如图 16-3 所示，Voronoi 多边形的边与裁剪线有 4 种关系：两点同在边界内、两点同在边界外、起点在边界内而终点在边界外、起点在边界外而终点在边界内。

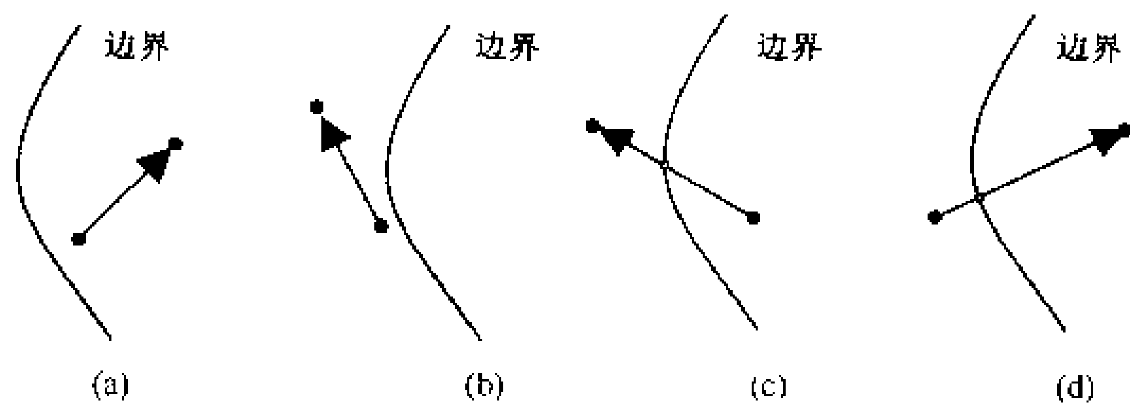


图 16-3　Voronoi 多边形裁剪原理

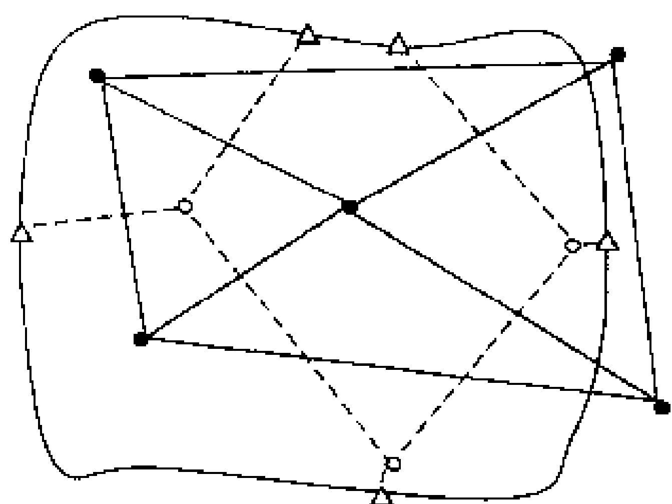


图 16-4　Voronoi 多边形裁剪结果

裁剪后以被裁剪边与裁剪边界线的交点作为分割点，对区域边界部分的离散点的 Voronoi 图重新划分，得到这些点的有效 Voronoi 区。如图 16-4 所示，为基于土地边界对图 16-2 裁剪后的平面点集有效 Voronoi 图。

4. 平面点集约束 Voronoi 图构建

与平面点集约束 D-TIN（CDT）对应的 Voronoi 图为平面点集约束 Voronoi 图。Okabe A. 等（1992）曾提到一种局部修正算法，该算法可以根据平面点集中的线状障碍物，将 OVD 图修正为可视最短路径 Voronoi 图。李武龙等（1998）在局部修正算法的基础上，提出了分段剖分、整体剖分和统一剖分的思路，以期提高剖分区重新分配的效率和保证剖分区重新分配的准确性。

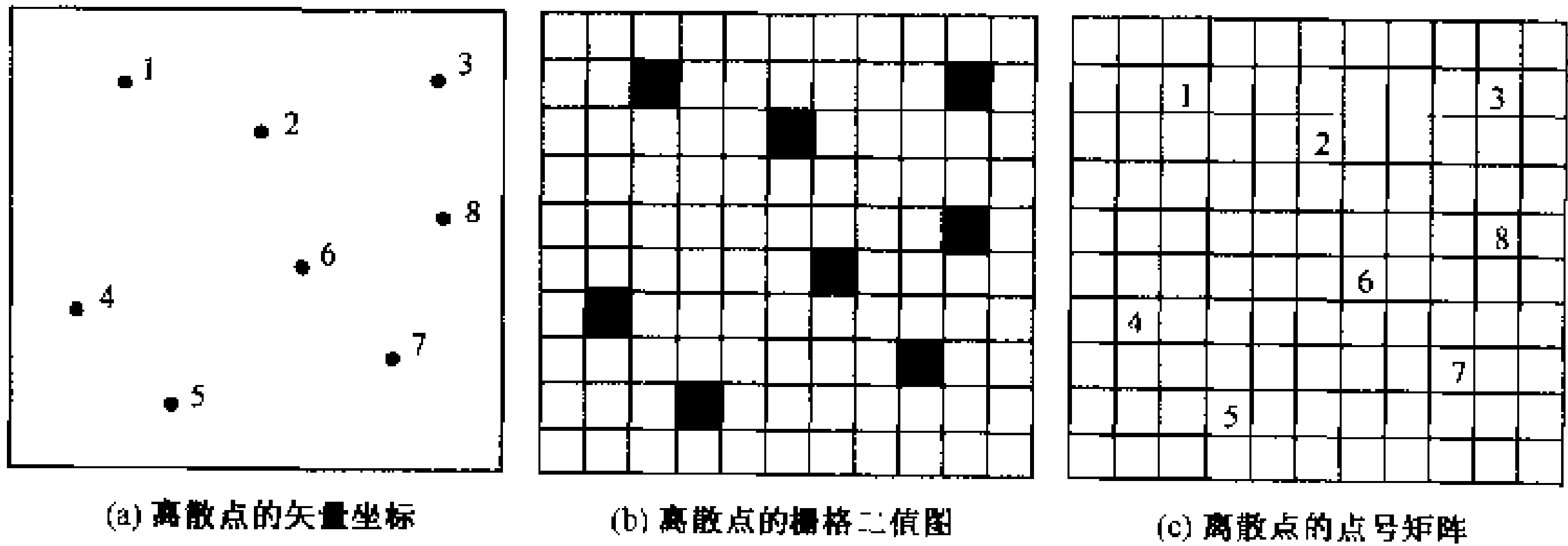
16.1.2　基于栅格的 Voronoi 图自动构建

如果将拟建立 TIN 的区域与一幅二值数字影像相对应，凡是与数据点对应的像素

灰度值为 1，其他像素的灰度值为零，则可以利用基于集合论的数学形态学（Mathematic Morphology）方法，进行形态变换来建立 Voronoi 图。

1. 矢量-栅格转换

通常，离散采样点首先是矢量数据。将离散点矢量数据转换为栅格数据的关键是栅格尺寸的选择。若栅格尺寸太大，将直接影响求 Voronoi 图的质量；若太小，则会增加存储量和计算时间。选择栅格尺寸应遵循的基本原则是：同一栅格中不允许有两个离散点，而且任意两个相邻的栅格中不允许同时有离散点存在。所以，通常选取离散点集中处距离最小的两点之间距离的 3 倍，作为栅格尺寸。转换后，离散点所在栅格的值取为离散点编号（多值化），空白栅格则置空。如图 16-5 所示，研究区被转化为一幅栅格二值影像图和一幅基于离散点编号的多值影像图。



(a) 离散点的矢量坐标　(b) 离散点的栅格二值图　(c) 离散点的点号矩阵

图 16-5　离散点集转换为栅格图

2. 产生 Voronoi 图

数学形态学中有两类基本操作：膨胀或扩张（Dilation）和腐蚀（Erosion），其中扩张可以用于基于栅格的 Voronoi 图生成。设 A 是原始影像，B 是结构元，$A \oplus B$ 是扩张的结果，图 16-6 所示为一个扩张实例：

$$\text{Dilation}: A \oplus B = \bigcup_{b \in B} A_b \tag{16-4}$$

以八邻域为结构元，在所获得的离散点二值图中，采用图像扩张技术对各离散点栅格向其八邻域方向进行扩张；每扩张一次，判断所扩张到的栅格是否已经被占用；如果是，则此栅格为本离散点栅格的 Voronoi 边界栅格，编号为 B；如果不是，则此栅格为本离散点栅格的 Voronoi 区内栅格，并继续扩张；直到所有栅格均被扩张到。

图 16-5 所示 8 个离散点的扩张过程见图 16-7。

根据编号为 B 的栅格，产生各点的 Voronoi 边界：此时要根据 B 为两点共用，还是三点或多点共用，以及 B 的走向，来综合决定 B 栅格内扩张边界的划分方式。此后，即可以得到离散点的矢量化 Voronoi 图。图 16-8 所示为扩张结果和 Voronoi 矢量图。

此外，也有学者提出用距离变换来实现离散空间的 Voronoi 图构建。距离变换的基本思想是：根据一定的距离定义方式生成二维空间中离散目标的距离图，其中每一点的

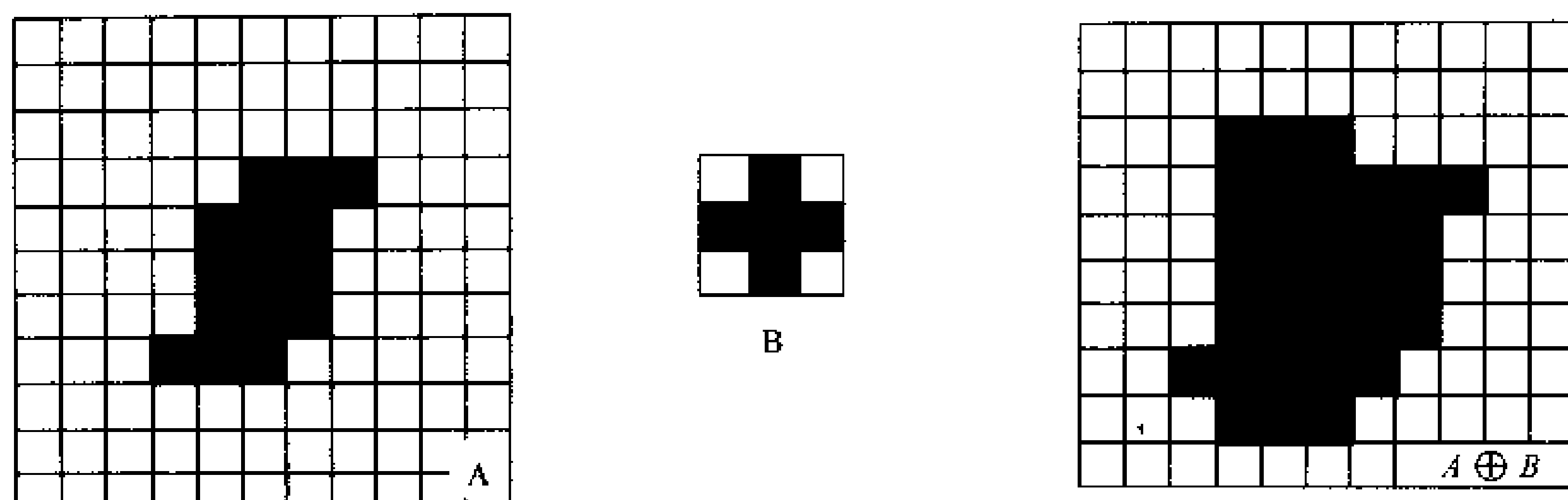


图 16-6 栅格扩张过程

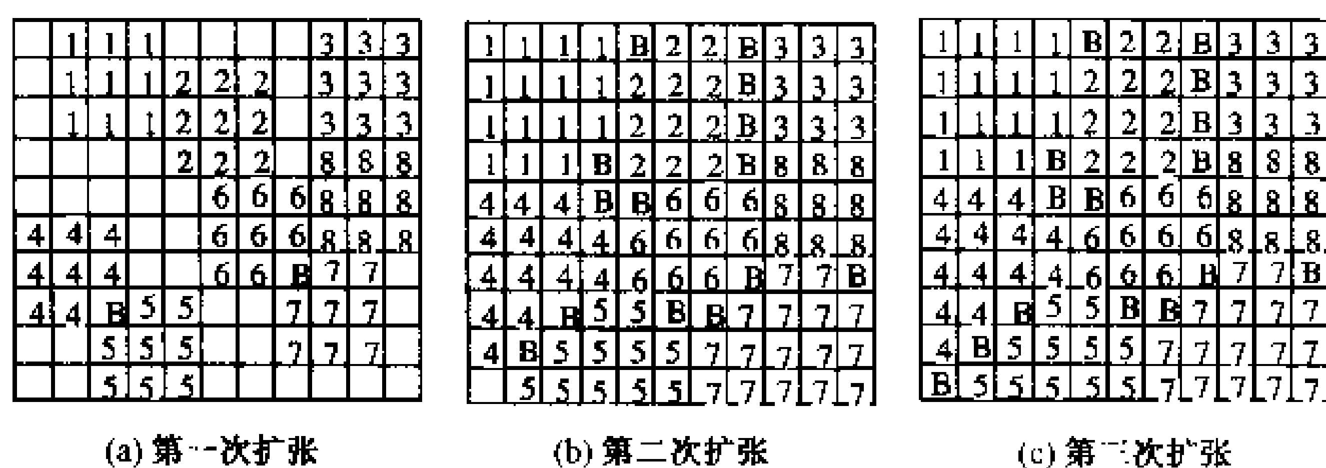


图 16-7 离散点栅格化扩张过程

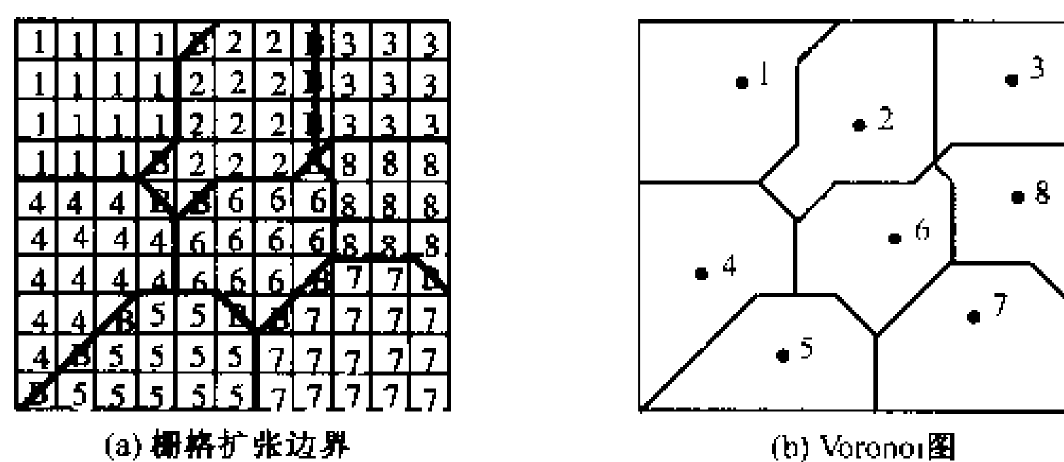


图 16-8 离散点扩张边界与 Voronoi 图

距离到所有空间目标中距离最小。如果某点到两个以上目标的距离相等，则该点为一个 Voronoi 边界点，否则隶属于某一目标。所有隶属于某一空间目标的点即构成该目标的 Voronoi 区；各目标的 Voronoi 区的集合即为二维空间离散目标集的 Voronoi 图。选择合适的距离定义方式，就可以将距离变换思想用于栅格空间。

通常，可以采用 3×3、5×5、7×7 模板的邻近像元的局部距离来整数近似地计算欧几里德距离。李成名等（2000）研究表明，这种处理方式使得所生成的 Voronoi 图的最大偏差总是与水平方向的距离成正比。鉴此，李成名等（2000）提出，可以根据欧几里德距离作用的圆条件要求，通过与目标特征匹配选择不同的模板（如矩形模板、菱形模板等），来进行动态距离的变换，使得误差始终保持在一个像素左右。

16.1.3 基于 Voronoi 图的 D-TIN 构建

平面点集 Voronoi 图和 D-TIN 可以互相转换，它们是一组对偶图。按照 Voronoi 图中多边形边界相邻的原则，将两两相邻的 Voronoi 图内离散点连接起来，就可以得到该平面点集的 D-TIN。这即是第 15 章中提及的基于 Voronoi 图的 Delauny TIN 构建算法的基本原理。以下分有、无约束条件进行讨论。

1．无约束 D-TIN

基于平面点集的无约束 D-TIN，可以在按上述算法生成如图 16-8（a）所示的 Voronoi 图上直接产生，如图 16-9 所示。

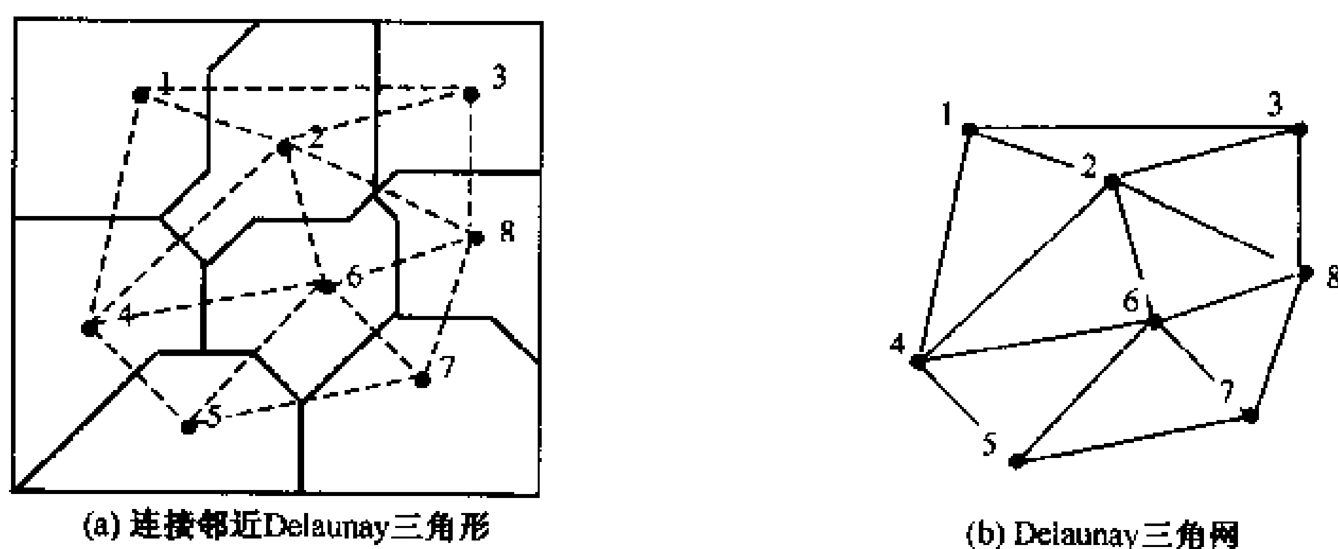


图 16-9　基于 Voronoi 图生成无约束 D-TIN 实例

2．CD-TIN

矢量环境下构建约束 TIN 是先建立离散点的 D-TIN，然后再考虑约束条件进行 TIN 的局部修改。与此相反，在栅格环境下构建约束 TIN，则是首先考虑约束条件，然后建立 TIN。其技术关键是对约束边进行栅格化和特殊编码。约束边的栅格化与一般矢量线、矢量面的栅格化过程一样，其过程如 16.2.2 节和 § 12.1 节所述。

对栅格化后的约束边给予特殊编码，其目的一方面是反映约束线段两点之间的关系，另一方面是在进行离散点扩张变换时起到防火墙的作用。除此之外，其他步骤与

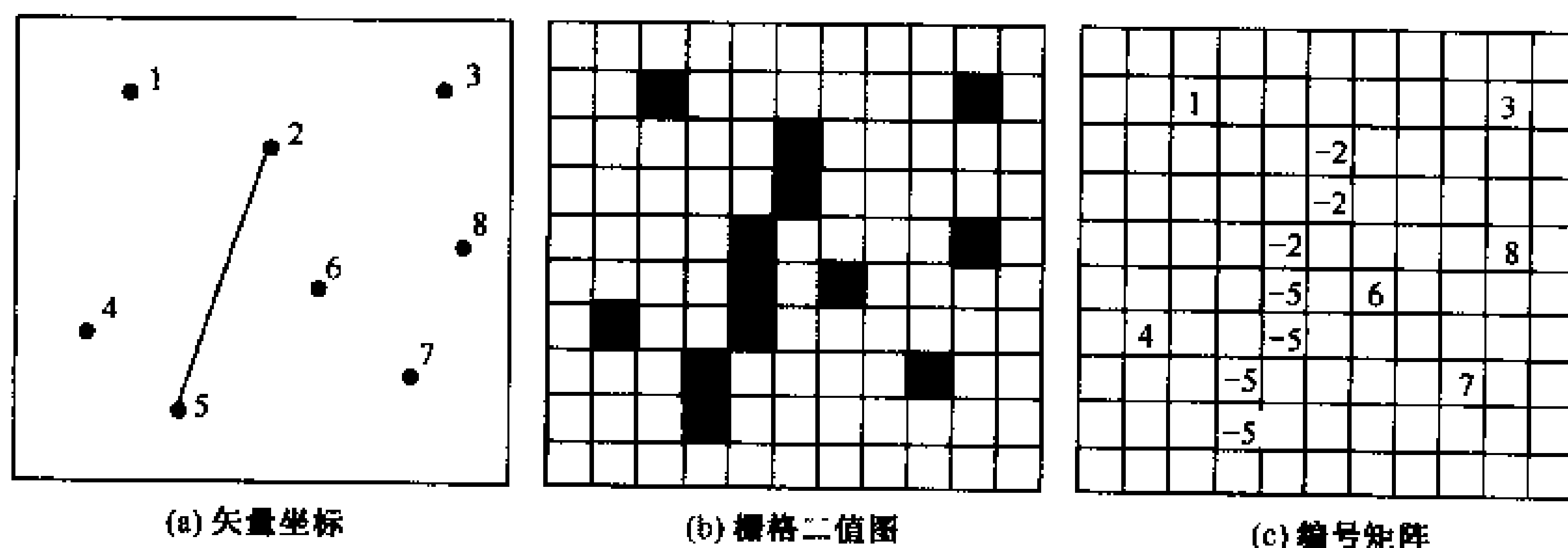
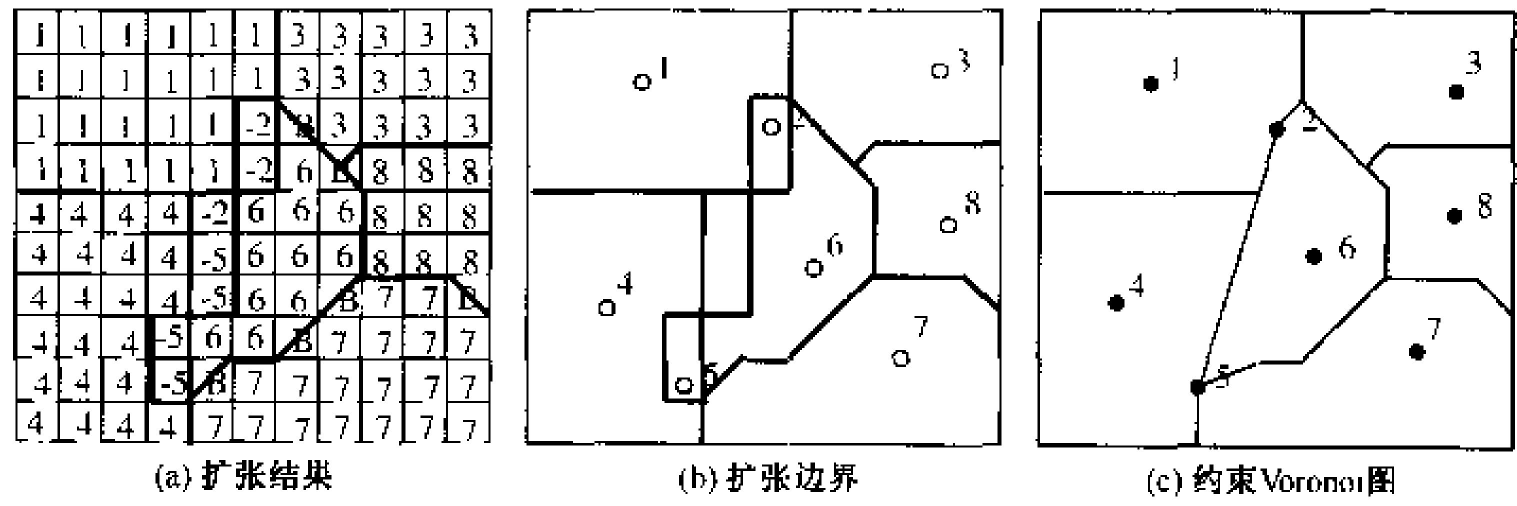


图 16-10　离散点与约束边矢量数据转换为栅格数据

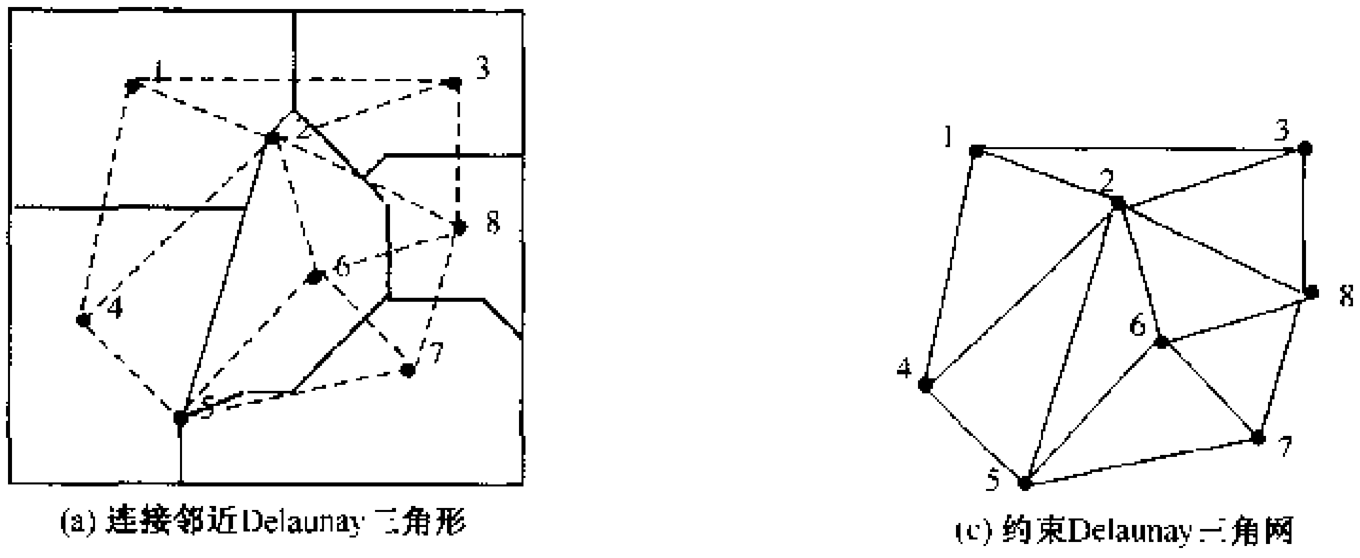
16.2.2 节的基本步骤相似。

以图 16-5 为例，若离散点 5、2 的连线是一条约束边，则其转化为栅格数据的结果如图 16-10 所示。经约束边界控制生成的 Voronoi 图如图 16-11 所示。在已经生成的约束 Voronoi 图基础上，根据多边形的拓扑关系，连接邻近多边形之间的点得到 Delaunay 三角形，进而得到区域带约束条件的 Delaunay 三角网，其结果如图 16-12 所示。



(a) 扩张结果　(b) 扩张边界　(c) 约束Voronoi图

图 16-11　离散点约束扩张边界与约束 Voronoi 图



(a) 连接邻近Delaunay三角形　(c) 约束Delaunay三角网

图 16-12　带约束 D-TIN

§ 16.2　线/面集 Voronoi 图构建算法

国外许多学者对平面特殊复杂实体的 Voronoi 图生成算法，如线段（Kirkpatrick D.，1979）、线状或非交圆片状（Lee D. et al, 1981）、任意圆片状（Sharir M.，1985）、平面凸壳（Leven D. and Sharir M.，1987）、平面曲线（Yap C.，1987）、平面单连通域（Lee D. et al.，1981；Fortune S.，1986；Chou J. J.，1995；Held M.，1998）、连续移动的点、线与凸壳（Tokuyama T.，1988；Aonuma H. et al.，1990；Roos T.，1991）等的 Voronoi 图生成算法做了深入研究。我国学者也对有线形障碍体的可视最短路径（李武龙等，1998）、基于栅格的面条目标（李成名等，2000）、任意平面域（刘晖等，1999）、平面多连通域（张大卫等，1997；付庄等，2000）等的 Voronoi 图生成算法及二维地理空间离散分布对象邻近关系模糊查询的意象式 Voronoi 模型（胡志勇等，2001）等做了研究和探索。

16.2.1 基于 C^1 偏置的 Voronoi 图构建算法

显然，角平分线、多边形内切圆圆心满足 Voronoi 图的定义。多年来，人们一直从这两种角度出发研究图形的线/面集合图形 G 的 Voronoi 图。以对角线方式寻找 Voronoi 图时，首先需要将 G 离散成近似多边形 G′，再通过对 G′中所有不平行边之间可能存在的角平分线、平行边之间可能存在的对称轴线求交，最终获得 G 的近似 Voronoi 图。这种方法在图形离散、直线求交时计算复杂；而以多边形内切圆圆心方式寻找 Voronoi 图时，完全是试探进行：在图形 G 的某一边界点变化直径取值，获得内切圆后沿 G 的边界滚动一定距离到达另一边界点，重复上述操作直到结束。这两种方法都是近似方法。

刘晖等（1999）基于 C^1 连续曲线的偏置算法，提出了一种获取 G 的 Voronoi 图的简单方法，G 的边界可以是任意形式的曲线，也可以是任意的连通域（任意个内部孔洞）。以下对这种方法作一介绍。

1．偏置曲线的特征

设 C^1 连续曲线 Γ 的参数表达为：$\Gamma(u)=(X(u),Y(u))$；$u\in[u_0,u_1]$；$u_0\leqslant u_1$。规定 $\Gamma(u)$沿 $u_0\to u_1$变化时为正向，并对应各点的切向矢量（以下简称切矢）。设各点切矢逆时针旋转后所得的矢量为各点的正向法矢（以下简称正法矢）。

所谓偏置曲线，就是对 $\Gamma(u)$上的每一点，均沿其正法矢偏移同一个量值，移动后的点集即构成原曲线 $\Gamma(u)$的等距偏置曲线 $\Gamma_{\text{off}}(u)$，其数学描述如下：

$$\begin{cases} X_{\text{off}}(u)=X(u)-\tau\dfrac{Y'(u)}{\sqrt{X'^2(u)+Y'^2(u)}} \\ Y_{\text{off}}(u)=Y(u)+\tau\dfrac{X'(u)}{\sqrt{X'^2(u)+Y'^2(u)}} \\ \Gamma_{\text{off}}(u)=[X_{\text{off}}(u),Y_{\text{off}}(u)];u\in[u_0,u_1];u_0\leqslant u_1;\tau\in R^2 \end{cases} \tag{16-5}$$

当 $\tau>0$ 时，$\Gamma_{\text{off}}(u)$ 为正向偏置曲线；当 $\tau<0$ 时，$\Gamma_{\text{off}}(u)$ 为负向偏置曲线。

对式（16-5）的前两项求导，得

$$\begin{cases} X'_{\text{off}}(u)=X'(u)\left[1-\tau\dfrac{Y''(u)X'(u)-Y'(u)X''(u)}{\sqrt[1.5]{X'^2(u)+Y'^2(u)}}\right] \\ Y'_{\text{off}}(u)=Y'(u)\left[1-\tau\dfrac{Y''(u)X'(u)-Y'(u)X''(u)}{\sqrt[1.5]{X'^2(u)+Y'^2(u)}}\right] \end{cases} \tag{16-6}$$

令

$$\bar{k}(u)=\tau\frac{Y''(u)X'(u)-Y'(u)X''(u)}{\sqrt[1.5]{X'^2(u)+Y'^2(u)}} \tag{16-7}$$

则式（16-6）简化为

$$\begin{cases} X'_{\text{off}}(u)=X'(u)[1-\tau\bar{k}(u)] \\ Y'_{\text{off}}(u)=Y'(u)[1-\tau\bar{k}(u)] \end{cases} \tag{16-8}$$

从式(16-7)可以看出：当 $\tau\bar{k}(u)<1$时，$\Gamma_{\text{off}}(u)$各点处的切矢与 $\Gamma(u)$对应点的切矢方向相同，仅是模发生了改变。由此可以推导出偏置曲线的以下性质：

性质1：给定 $\Gamma(u)$，对其进行 n 次正向偏置，各次偏置量分别为 τ_1、τ_2、…、τ_n。若 $\tau = \sum_{i=1}^{n}\tau_i$ 满足 $\tau \bar{k}(u) < 1$，则这 n 次正向偏置的结果与偏置量为 $\tau = \sum_{i=1}^{n}\tau_i$ 的一次偏置的结果相同，即偏置无累积误差。

性质2：$|\bar{k}(u)|$ 为 $\Gamma(u)$ 对应各点的曲率，当 $\bar{k}(u)>0$ 时，对应点的正法矢指向该点的曲率中心；当 $\bar{k}(u)<0$ 时，对应点的正法矢背离该点的曲率中心；当 $\bar{k}(u)=0$ 时，正法矢不存在，可以将 $\Gamma(u)$ 在对应点分开，当作多条 C^1 曲线分别处理。

性质3：给定 $\Gamma(u)$，设当 $u=u_{max}$ 时 $\bar{k}(u_{max})$ 最大且 $\bar{k}(u_{max})>0$，取 $\tau=1/\bar{k}(u_{max})$ 进行正向偏置，所得 $\Gamma_{off}(u)$ 在 $\Gamma(u_{max})$ 对应曲率处分成两条 C^1 连续曲线；设当 $u=u_{min}$ 时 $\bar{k}(u_{min})$ 最小且 $\bar{k}(u_{max})<0$，取 $\tau=1/\bar{k}(u_{min})$ 进行正向偏置(此时偏置量为负值)，结论相同。此时，所得 $\Gamma_{off}(u)$ 比 $\Gamma(u)$ 少一个元素，$u=u_{max}$ 和 $u=u_{min}$ 时对应的点不再属于 $\Gamma_{off}(u)$。

性质4：给定 $\Gamma(u)$，对其进行一次正向 τ 偏置后，$\Gamma_{off}(u)$ 上各点与 $\Gamma(u)$ 对应点满足：

$$k_{off}(u) = \frac{\bar{k}(u)}{1-\tau\bar{k}(u)} \tag{16-9}$$

性质5：对于单独存在的孤点，C^1 及 $\bar{k}(u)$ 无定义。此时，可 $k(u)=0$，其正法矢方向及参数区间在 $0\sim2\pi$ 中根据需要任意选取。此时，对孤点进行正向 τ 偏置后得到以该点为圆心、以 τ 为半径的圆或一段圆弧。

2. Voronoi 图生成方法

假设在一定精度范围内，可以用一系列的 C^1 连续曲线 $\Gamma_i(u_i)=[X_i(u_i), Y_i(u_i)]^T$（其中 $u_i\in[u_{i,low}, u_{i,up}]$），来逼近图形 G 的任何边界(包括孔洞)。这组曲线可以分为有向相邻(首尾相连)和有向不相邻两类，通称为有向相关。规定有向相邻曲线连接处的夹角为前一曲线正法矢逆时针旋转到后一曲线正法矢的角度(取值 $0\sim2\pi$)。

表 16-1 所示为曲线偏置的 6 类典型条件和偏置处理方法。图 16-13 至图 16-18 为对应的曲线偏置示例。

表 16-1　曲线偏置的分类（据刘　晖，1999，整理）

分类	条件	偏置处理方法
Ⅰ	有向相邻，正法矢夹角不大于 π	前段曲线的参数上限减小、后段曲线的参数下限增加
Ⅱ	有向相邻，正法矢夹角大于 π	前段曲线的参数上限增加，后段曲线的参数下限减小
Ⅲ	有向不邻，偏置曲线不相切	无满足 Voronoi 图定义的点，不作处理
Ⅳ	有向不邻，偏置曲线相切，且切点为某一偏置曲线的起点或终点	若为起点，则减少另一偏置曲线的上限；若为终点，则增加另一偏置曲线的下限
Ⅴ	有向不邻，偏置曲线相切，目切点非以上情况	在切点处将两条偏置曲线各自分开，得到两组相邻偏置曲线
Ⅵ	有向不邻，曲线平行	偏置曲线可能重合，切点变为切线，整条切线均满足 Voronoi 的定义

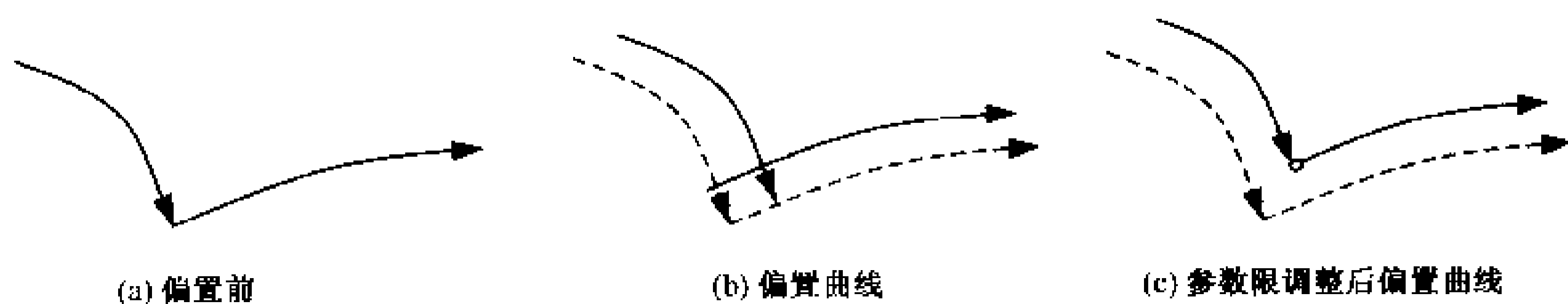


图 16-13　第Ⅰ类偏置结果

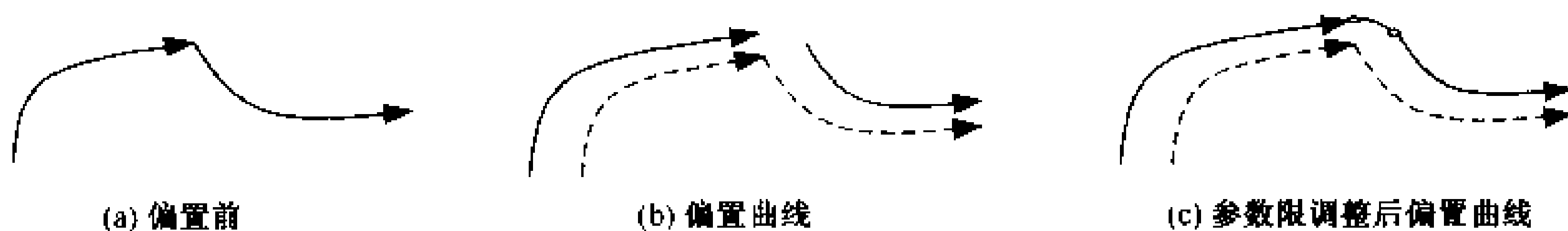


图 16-14　第Ⅱ类偏置结果

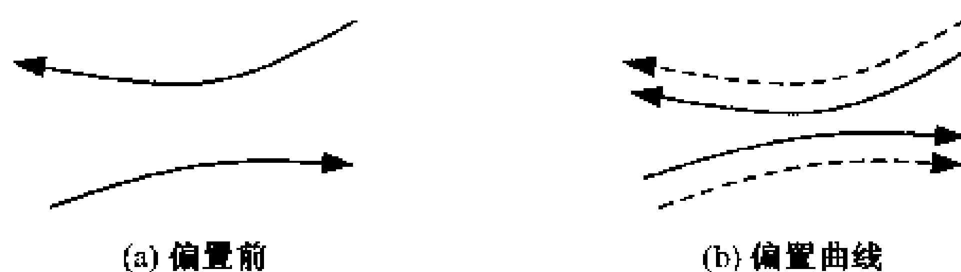


图 16-15　第Ⅲ类偏置结果

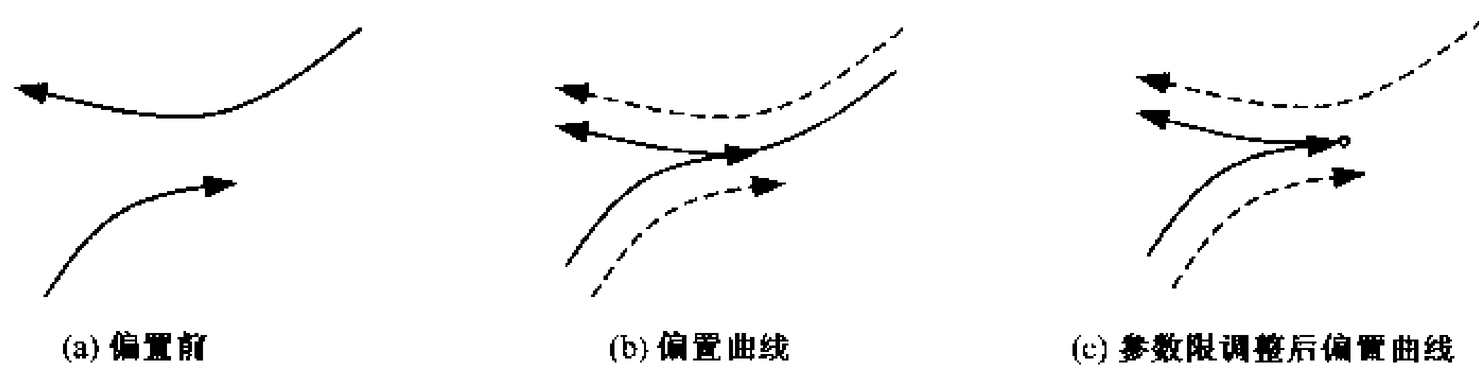


图 16-16　第Ⅳ类偏置结果

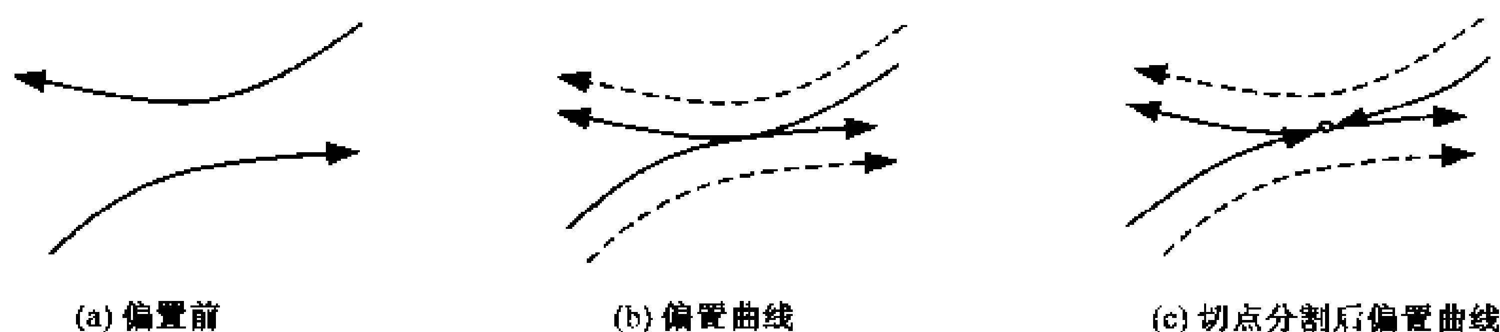


图 16-17　第Ⅴ类偏置结果

综上所述，给定一组 C^1 连续曲线，在不断偏置过程中总会得到一系列有向相邻点（包括交点、切点）和切线，可以证明这就是该组 C^1 连续曲线满足 Voronoi 图定义的元

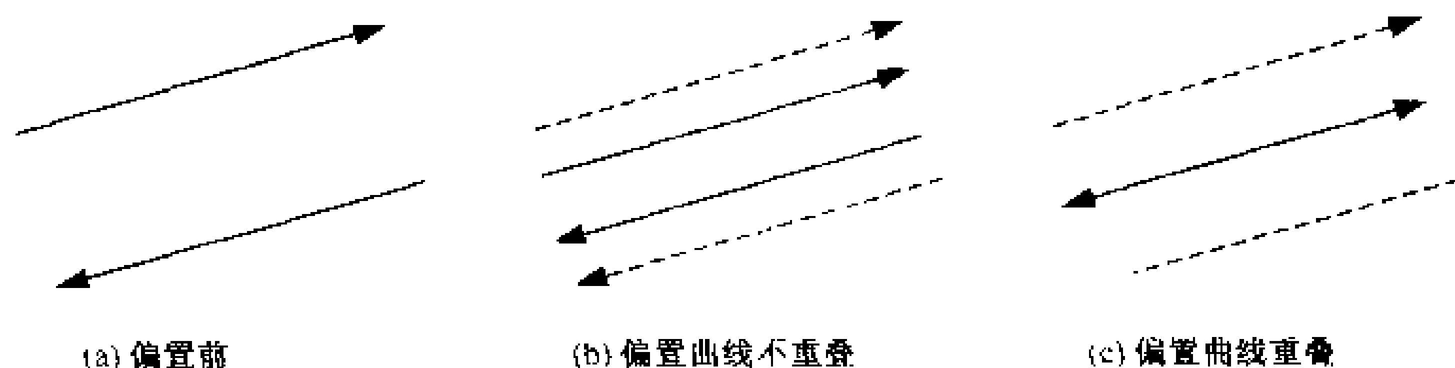


图 16-18　第Ⅵ类偏置结果

素。因此，推求图形 G 的 Voronoi 图转换为两个问题：

(1) 首先将 G 用一组 C^1 连续曲线表示；

(2) 其次对这组 C^1 连续曲线不断进行偏置得到一系列有向相邻点，将这些点依次相连，即得到 G 的 Voronoi 图。

该方法的关键是要正确确定曲线的方向，外部边界曲线沿逆时针方向为正向，内部孔洞边界曲线沿顺时针方向为正向。当偏置量 τ 取正值且沿曲线正法矢偏置时，得到 G 的内部骨架或内部 Voronoi 图；当偏置量 τ 取负值且沿曲线正法矢偏置时，得到 G 的外部骨架或外部 Voronoi 图。

刘晖等（1999）基于这一算法使用 AutoCAD 技术实现了单连通域边界分别为直线和曲线、图形分别包含一个和两个孔洞的图形 G 的 Voronoi 图的自动生成。

16.2.2　基于动态距离变换的 Voronoi 图构建算法

所谓面条目标，是指占据一定空间的点、线、面和复杂目标的总称（李成名等，2000)。为了基于动态距离变换研究面条目标的 Voronoi 图的生成，李成名等对面条目标之间的距离和面条目标的 Voronoi 图进行了以下定义：

定义 1：设 x，y 是两个 n 维的点，其坐标为 $(x_1, x_2, \cdots, x_n)$ 和 $(y_1, y_2, \cdots, y_n)$，其间的距离为：

$$\text{Distance} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \tag{16-10}$$

定义 2：设 C 是线状元素，x 是一个点，其间的距离为：

$$\text{Distance}(x, C) = \text{Min}\{\text{Distance}(x, y), y \in C\} \tag{16-11}$$

定义 3：设 C_1、C_2 是线状元素，x、y 是两个点，C_1、C_2 之间的距离为：

$$\text{Distance}(C_1, C_2) = \text{Min}\{\text{Distance}(x, y), x \in C_1, y \in C_2\} \tag{16-12}$$

定义 4：设 R 是面状元素，x 是一个点，其间的距离为：

$$\text{Distance}(x, R) = \text{Min}\{\text{Distance}(x, y), y \in R\} \tag{16-13}$$

定义 5：设 R_1、R_2 是面状元素，x、y 是两个点，R_1、R_2 之间的距离为：

$$\text{Distance}(R_1, R_2) = \text{Min}\{\text{Distance}(x, y), x \in R_1, y \in R_2\} \tag{16-14}$$

定义 6：设 C、R 分别是线状元素和面状元素，x、y 是两个点，C、R 之间的距离为：

$$\text{Distance}(C, R) = \text{Min}\{\text{Distance}(x, y), x \in C, y \in R\} \tag{16-15}$$

定义 7：设一幅图中包含 n 个面条目标元素 P_1，P_2，…，P_n，目标 P_i（$i=1\sim n$）Voronoi 多边形定义为：

$$\text{Voronoi}(P_i) = \{\text{Distance}(x, P_i) < \text{Distance}(x, P_j), j = 1 \sim n, j \neq i\} \quad (16\text{-}16)$$

P_i（$i=1\sim n$）Voronoi 多边形的边界为：

$$B(P_i) = \{\text{Distance}(x, P_i) = \text{Distance}(x, P_j), j = 1 \sim n, j \neq i\} \quad (16\text{-}17)$$

则图中面条目标集合的 Voronoi 图为：

$$\text{VD}(P) = \left\{\bigcup_{k=1}^{n} \text{Voronoi}(P_k)\right\} \cup \left\{\bigcup_{k=1}^{n} B(P_k)\right\} \quad (16\text{-}18)$$

基于以上定义，李成名等在所提出的动态距离变换思路的基础上，对面条目标的 Voronoi 图的生成和 Voronoi 图邻接关系的获取问题进行了研究。

由于线集、面集矢量算法的复杂性，由线、面元素组成的复杂实体的 Voronoi 图的构建算法的研究更为困难。为便于简化，Yang 和 Gold（1996）曾提出一种“点-线”模型，把复杂实体分解成点和直线，先构建点线的 Voronoi 图，再转换为复杂实体的 Voronoi 图。Yang（1998）进一步对地图目标动态 Voronoi 图（VMO）的模型进行了研究，并具体应用于森林数据管理。

此外，Aurenhammer F. 和 Edelsbrunner H.（1984）提出了构建倍增的加权 Voronoi 图算法，基于矢量实现了点的加权 Voronoi 图；张有会（1995）则将点的加权 Voronoi 图推广到线段加权 Voronoi 图，研究了两线段间 Voronoi 边的具体形状和画法，并证明两线段间的 Voronoi 边是二次曲线；王新生等（2002）则将线的加权 Voronoi 图和面的加权 Voronoi 图应用于界定城市空间影响范围，并实现了基于栅格方法的若干类 Voronoi 图的自动构建。限于篇幅，这里不能详细介绍，有兴趣的读者可以参阅有关文献。

§16.3 球面 Voronoi 图构建算法

近年，人们将研究区域逐渐从局部范围扩展到全球范围，导致一个新的热点研究领域即球面 Voronoi 图的出现。加拿大的 Lukatela H.（1987；1989）在其主持研制的 HIPPARCHUS 系统中利用球面 Voronoi 剖分，建立了地表、航空和卫星运动的时空关系模型；澳大利亚的 Watson D. F.（1988；1994；1998）也对球面 Voronoi 图模型进行了研究，在其开发的 MODEMAP 系统中，研究了在球面空间利用 Voronoi 图进行球面距离和面积的计算、自然邻近插值、三角划分、密度估计等问题。

球面 Voronoi 图的研究对于全球数据的动态管理和球面空间关系的推理有重要意义。目前对球面 Voronoi 图生成算法的研究相对比较少，且均是基于球面点集的矢量算法，代表性的有 Aggenbaum（1985）的插入法［时间复杂度为 $O(n^2)$］和 Robert（1997）的分治法［时间复杂度为 $O(n\lg n)$］。Gold 和 Mostafavi（2000）试图把“点-线”模型扩展到球面上。此方法的优点是把复杂实体转换为简单要素，能够计算出球面复杂实体的矢量 Voronoi 图和处理球面实体的动态变化，但是此方法需要经过“分解”、“计算”、“合成”等步骤，结构复杂且算法繁琐，特别是缺乏数据的层次表达，难以从根本上解决球面海量数据的多层次综合处理和空间复杂分析。

16.3.1 球面点集 Voronoi 图构建算法

球面点集 Voronoi 图的相关定义如下：

定义 1：（球面大圆距离，Okabe A. et al.，1992）

球面 S 上任意两点 p_i、p_j 的矢量位置为 X_i、X_j，其间用互不相交的大圆弧 p_ip_j（其中心点为球心，该圆弧为两段圆弧中的较短者）连接，大圆弧 p_ip_j 的长度即为球面两点之间的距离：

$$d(p_i, p_j) = 2R\arccos(X_i^t X_j) \leqslant 2R\pi \tag{16-19}$$

定义 2：（球面点集 Voronoi 图，Okabe A. et al.，1992）

球面 S 上点集 $P = \{p_1, p_2, p_3, \cdots, p_n\}$ $(2 \leqslant n \leqslant \infty)$，$d(p_i, p_j)$ 为点 p_i 和 p_j 之间的球面距离，称：

$$V(p_i) = \{p \mid d(p, p_i) \leqslant d(p, p_j), i \neq j; i, j \in I_n; p \in S\} \tag{16-20}$$

为关于 p_i 的球面 Voronoi 多边形区域，称球面 Voronoi 多边形的集合为球面 S 上点集 P 的球面 Voronoi 图（如图 16 19）。

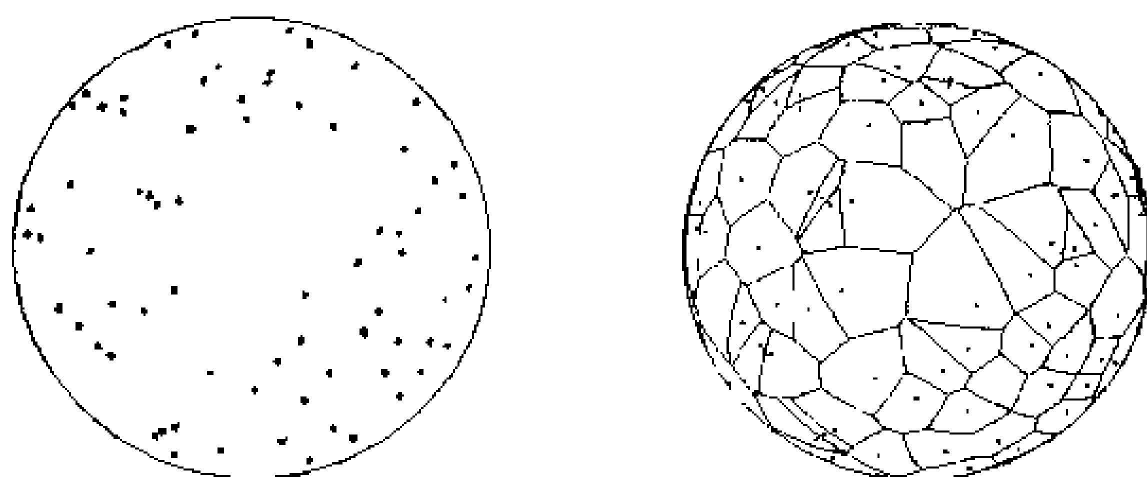


图 16-19　球面点集及其 Voronoi 图

（Geyer，2000）

徐寅峰等（1995）则将球面点集 Voronoi 图定义为：大圆弧 p_ip_j 的平分线 $\{x \in S \mid |xp_i| = |xp_j|\}$ 相交形成 $(n+1)$ 个球面凸多边形区域 $V(p_i)$，这 $(n+1)$ 个球面凸多边形区域将球面划分而成的凸网就称为球面点集 P 的 Voronoi 图，记为 Vor（P）。

上面提到的“球面多边形”是通过球面 S 上的点集定义的，就像平面上的凸多边形。我们定义“凸球面多边形”为：若通过球面多边形任意两点的大圆的小弧段嵌入在球面多边形的内部，此球面多边形称为凸球面多边形。

球面点集的 Voronoi 多边形具有以下性质（Okabe A. et al.，1992）：

性质 1：由公式（16-20）定义的集合是一个惟一的、有限的非空凸球面多边形。

性质 2：一个球面 Voronoi 多边形的边是一个大圆的弧。

性质 3：由 $P = \{p_1, p_2, \cdots, p_n\}$ $(n > 2)$ 产生的球面 Voronoi 图，设 n_v 和 n_e 为 Voronoi 图的顶点和边的数目，下面的公式成立：

$$n_v - n_e + n = 2 \tag{16-21}$$

定义 3：球面任意实体间的距离

设 P_1，P_2 为球面 S 上的两个子集，p_1 和 p_2 为 S 上的点，且 $p_1 \in P_1$ 和 $p_2 \in P_2$。集合 P_1 和 P_2 之间的距离定义为：

$$D(P_1, P_2) = \{d(p_1, p_2) \mid d(p_1, p_2) \leqslant d(p_i, p_j); p_i \in P_1, p_j \in P_2; i, j \in I_n\} \tag{16-22}$$

称此距离为集合 P_1 和 P_2 之间的球面距离。

定义 4：球面任意实体的 Voronoi 图

设 $P = \{P_1, P_2, \cdots, P_n\}$（$2 \leqslant n < \infty$）为球面 S 上的任意子集，$D(P_i, P_j)$ 为集合 P_i 和 P_j 之间的球面距离，称：

$$V(P_i) = \{P \mid D(P, P_i) \leqslant D(P, P_j), i \neq j, j \in I_n, P \in S\} \tag{16-23}$$

为关于 P_i 的球面 Voronoi 多边形区域，称球面 Voronoi 多边形的集合为球面 S 上任意实体（子集）P 的球面 Voronoi 图。

16.3.2　球面格网 Voronoi 图构建算法

为提高球面三角形地址编码与坐标转换效率，Dutton G.(1991)在正八面体的基础上，利用 ZOT 投影建立了 QTM（quaternary triangular mesh）多层次空间数据模型。赵学胜则选取球内接正八面体作为球面格网划分的基础，并采用"经纬度平分法"进行球面三角形的细分。同时，赵学胜基于 Voronoi 图栅格算法的整数近似原理（Borgefors G.，1986，1994）认为，在球面栅格空间也可以用邻近像元的局部距离近似代替"大圆距离"。由于在球面三角格网剖分中，每一个格网点与其相邻点组成近似正六边形，所以在球面格网距离变换中，每一个三角形格网是按六边形向外扩张的，如图 16-20。这即是"最大圆距离"在球面栅格空间的一个整数近似。

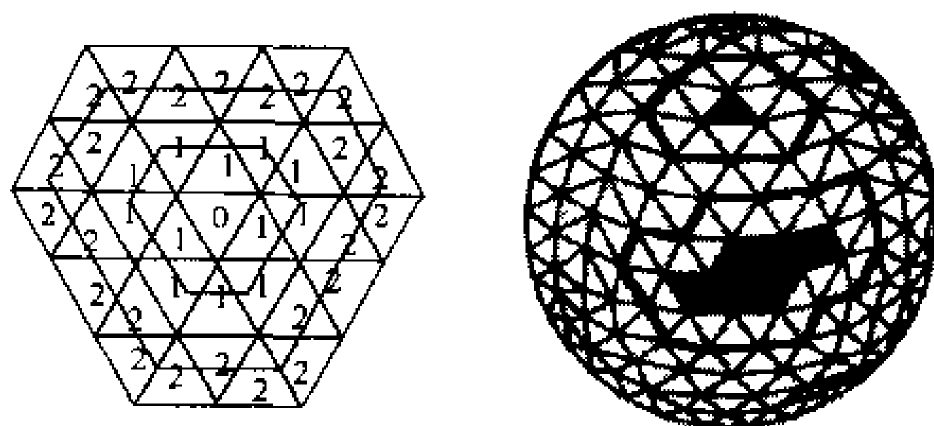


图 16-20　球面格网六边形模版和距离扩张
（赵学胜，2002）

为研究球面格网 Voronoi 图的生成算法，赵学胜（2002）基于数学形态学分别定义了球面三角格网膨胀和侵蚀操作如下：设 A 为球面三角格网的原区，B 为基本算子，则：

$$\text{扩张}\quad A \oplus B = \bigcup b \in BA_b$$
$$\text{侵蚀}\quad A \Theta B = \bigcap b \in BA_b \tag{16-24}$$

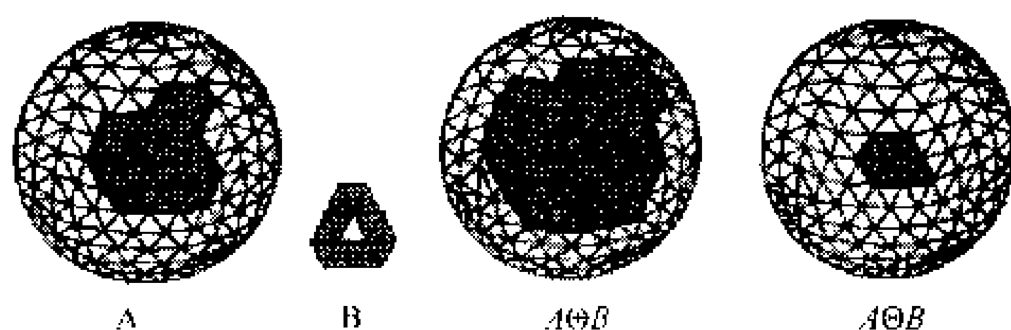


图 16-21 球面三角格网的膨胀和侵蚀操作

（赵学胜，2002）

1．单层次球面格网的 Voronoi 图（赵学胜，2002）

球面三角格网的 Voronoi 图可以按上面定义的球面三角形的膨胀原理生成。首先通过球面三角形的邻近搜索算法，确定出包围点、弧和曲面的边邻近和角邻近三角形，即搜索出区域内三角形的全部邻近三角形。然后剔除重复三角形，点、弧和曲面就生成一个膨胀区域，其边界与原始的点、弧和曲面的边界具有相同距离；重复进行此过程……，直到膨胀的区域与其他实体的膨胀区域相交，把相交的三角形格网连成线，即为两实体的 Voronoi 边。算法的程序框图如图 16-22 所示：

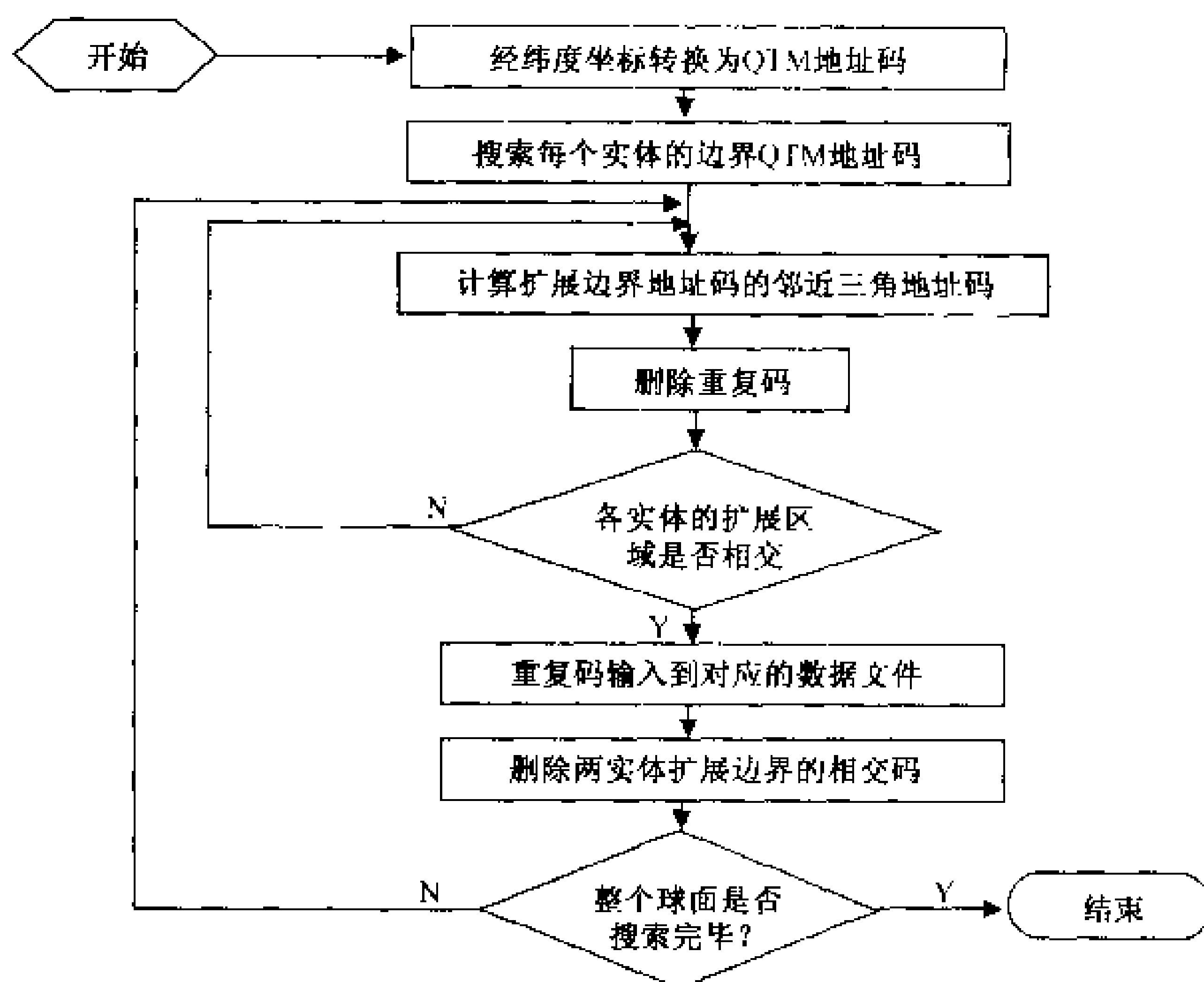


图 16-22 球面格网 Voronoi 图的生成算法框图

2．多层次球面格网的 Voronoi 图（赵学胜，2002）

由于球面实体的邻近结构也是一种空间拓扑关系，有时从低分辨率（粗格网）和高分辨率（细格网）得到的结果是一致的。因而，在不影响邻近关系的前提下，多层次递归生成具有邻近关系的球面 Voronoi 图，可以大大降低系统的时间和空间消耗。首先从

实体格网的低层次开始进行扩展，格网比较大；然后寻找不定点，在不定点邻近区域重新进行下一层次的实体扩展，以此类推……，直到没有不定点。这样就完成了整个区域的实体扩展，最后应用三角形的扩展值进行 Voronoi 边和结点的搜索：即具有两个值的三角格网是 Voronoi 边；具有三个（或以上）值的三角格网是结点。

以一个具体的例子说明多层次球面 Voronoi 图的生成算法原理：应用非层次算法，在层次 L 中，通过距离扩张可以得到实体的 Voronoi 图和邻近结构（如图 16-23）。同样在层次算法中，首先从层次 $L-1$ 进行距离变换［如图 16-24（a）］，存在一个不定点，在不定点的邻近区域（一个格网）进行下一层次（L）距离扩展［如图 16-24（b）］，以此类推……，直到没有不定点，完成全部实体的邻近关系构建。由于三角形格网是四叉树结构，所以每细化一个层次，三角形的个数增加 3 倍，而系统的花费是与三角形的个数成正比的。所以，在满足同样精度下，多层次扩展搜索的空间和时间消耗比单层次节省得多。

不定点的确定是层次 Voronoi 图生成算法的关键。只有存在不定点，才有必要在不定点的邻近区域进行下一层次的实体扩张。所以确定不定点位置和邻近区域构成了层次算法的主要研究内容。

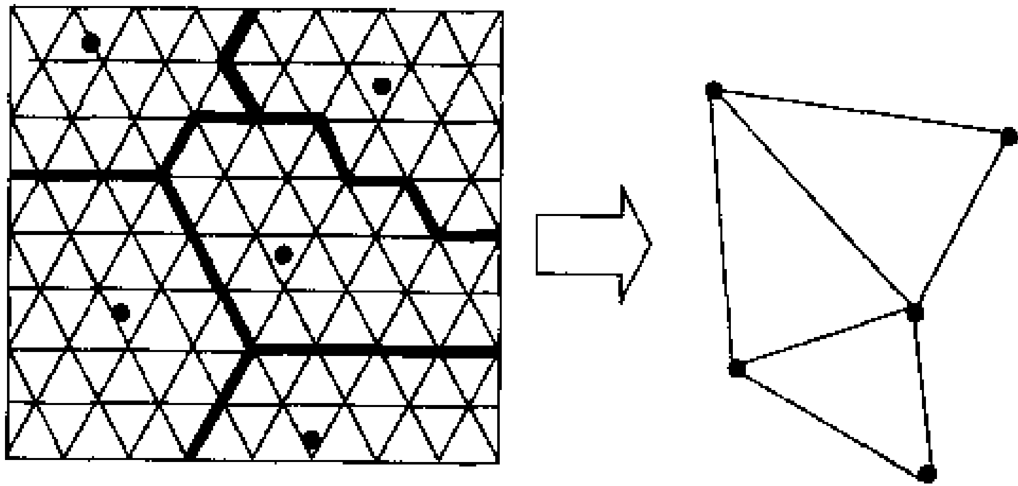


图 16-23　在层次 L 生成的 Voronoi 图和邻近结构

不定点的组成：

- 临界点——具有 4 个以上（包括 4 个）不同值的三角形点。
- 相邻结点——具有 3 个值的相邻三角形结点。

在具体处理中，每个临界点都需要进行下一层次的划分和重新扩展；但是对于相邻结点，则情况要复杂得多。但绝大部分都不需要下一层次的细分。在图 16-25 的相邻结点中，设四个实体特征点为 A，B，C，D，若特征点 A 在层次 $L+1$ 中的三角形与在层次 L 中的扩展三角形相邻，则其值为 1，否则为 0；同样定义特征点 B、C、D 的值。在总共 16 种情况下（如表 16-2），只有两种情况 {1，1，0，0} 和 {0，0，1，1} 需要下一层次细分。

基于以上原理，赵学胜利用 Visual C++ 语言在 OpenGL 三维可视化平台上开发了相应的实验程序。利用 QTM 地址码的邻近搜索，并根据球面实体（生长点或体）的并行膨胀原则，分别计算了球面点集、弧线集、面集和球面任意实体集的格网 Voronoi 图及其矢量化后的 Voronoi 图，其屏幕输出结果分别如图 16-26 所示。

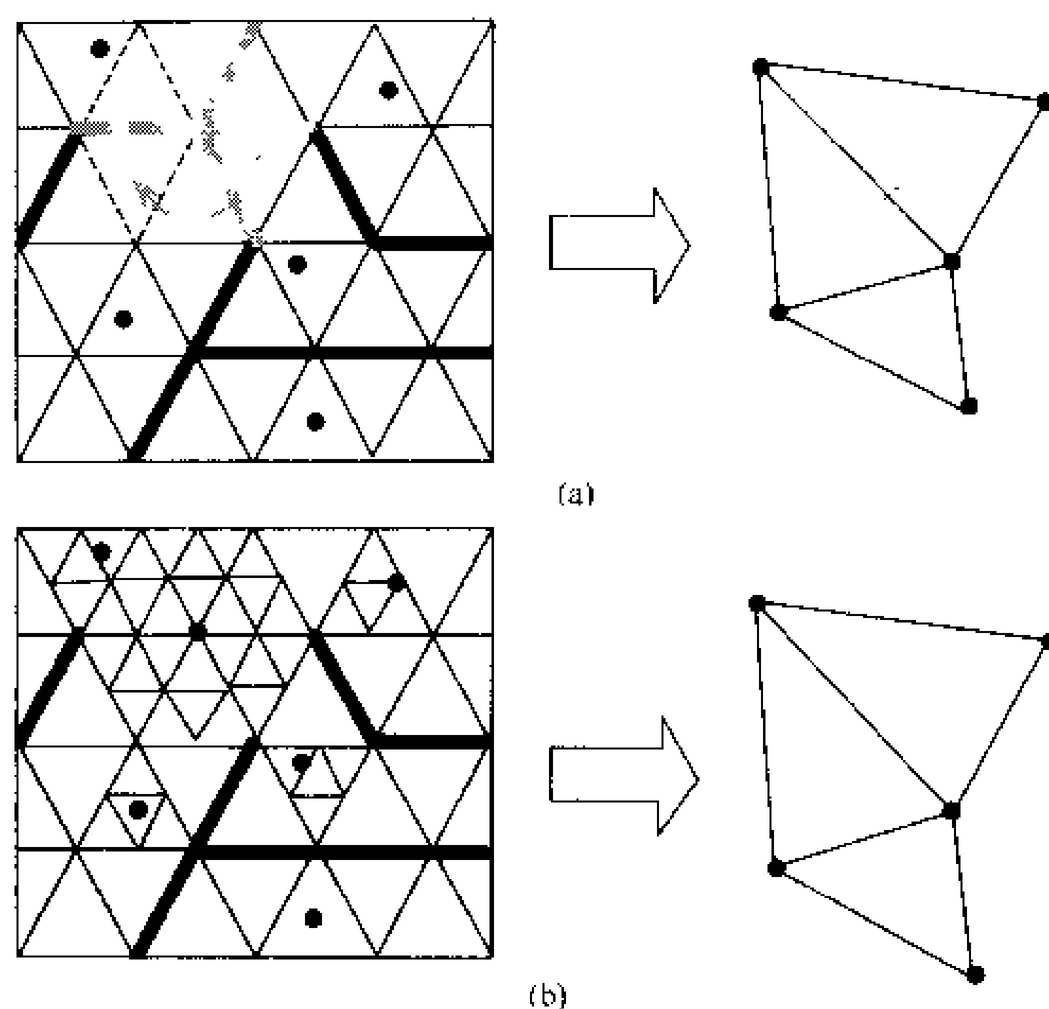


图 16-24　在层次 L-1 生成的 Voronoi 图和不定点区域的细化

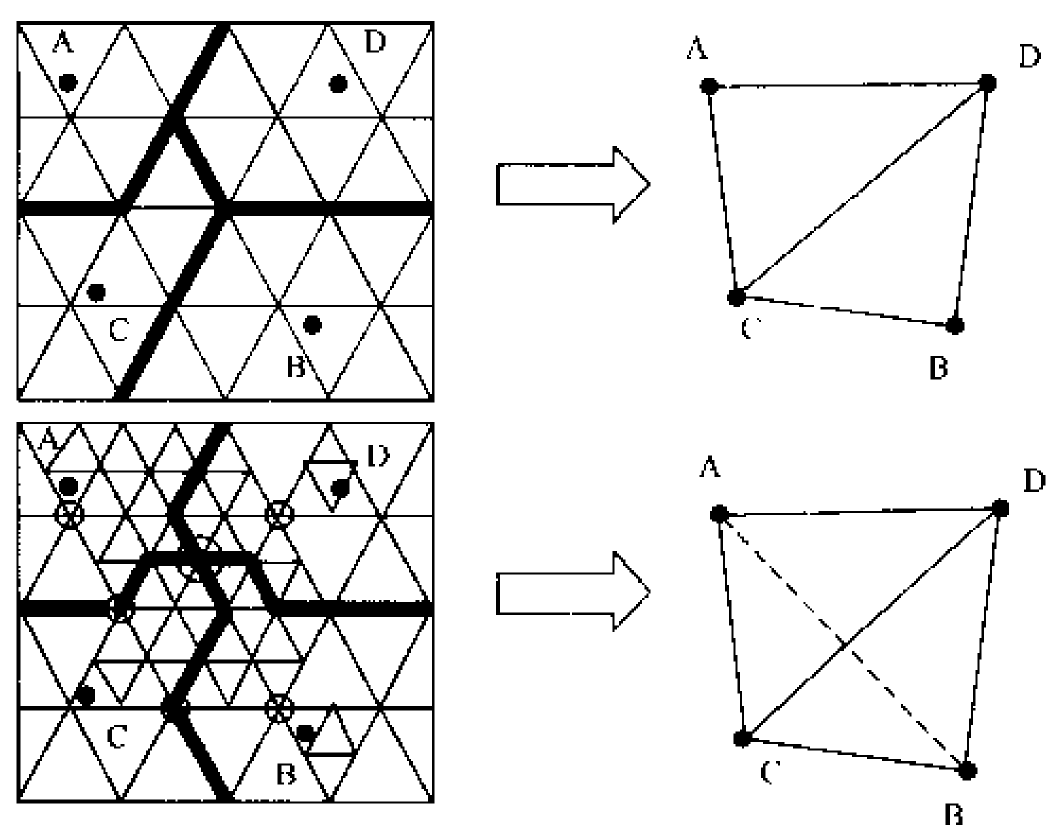
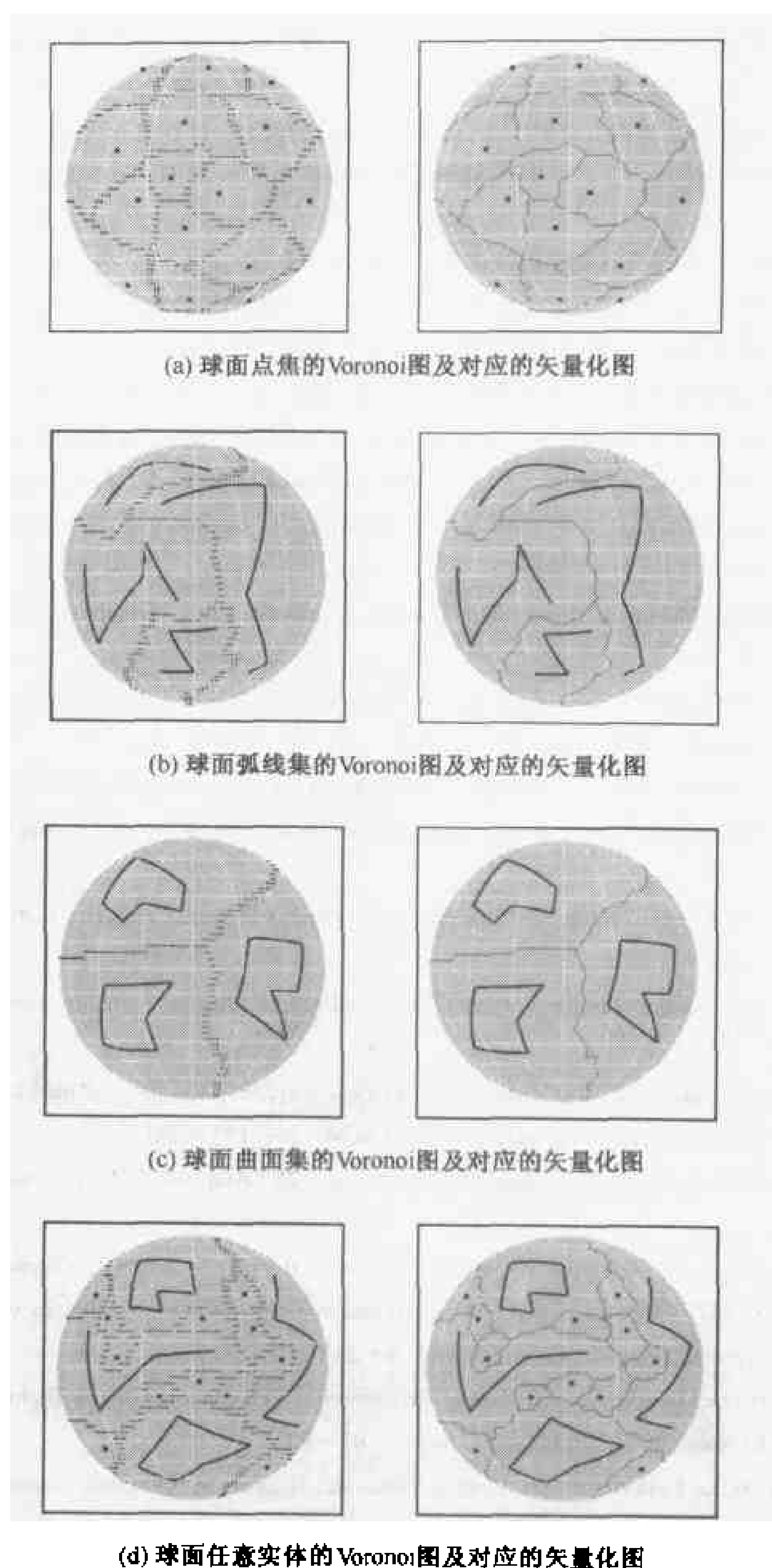


图 16-25　特殊相邻结点

表 16-2　特殊相邻结点的分类与处理

A, B, C, D	
{0, 1, 0, 0}	{0, 0, 0, 0}
{0, 1, 0, 1}	{0, 0, 0, 1}
{0, 1, 1, 0}	{0, 0, 1, 0}
{0, 1, 1, 1}	{0, 0, 1, 1}
{1, 1, 0, 0}	{1, 0, 0, 0}
{1, 1, 0, 1}	{1, 0, 0, 1}
{1, 1, 1, 0}	{1, 0, 1, 0}
{1, 1, 1, 1}	{1, 0, 1, 1}



(a) 球面点集的Voronoi图及对应的矢量化图

(b) 球面弧线集的Voronoi图及对应的矢量化图

(c) 球面曲面集的Voronoi图及对应的矢量化图

(d) 球面任意实体的Voronoi图及对应的矢量化图

图 16-26　球面 Voronoi 图生成结果

参 考 文 献

陈　军. Voronoi 动态空间数据模型. 北京：测绘出版社. 2002. 24－62

付　庄等. 2000. 多连通域 Voronoi 图生成算法的研究. 系统工程与电子技术. 22 (11)：88～90

胡志勇，何建邦，柯正谊. 2001. 地理空间意像模式的 Voronoi 模型. 地理科学. 21 (2)：177～182

李成名，陈　军. 1998. Voronoi 生成的栅格算法. 武汉测绘科技大学学报. 23 (3)：208～210

李成名，陈　军　2000．面条目标 Voronoi 图生成的动态距离变换策略　遥感信息．(1)：6－11
李武龙，陈　军．1998．线状障碍物的可视最短路径 Voronoi 生成　武汉测绘科技大学学报．23（2）：132～136
刘　晖，段宝岩．1999．任意平面区域的 Voronoi 图．西安电子科技大学学报　26（1）：118～123
刘卫东，唐泽胜．1995．三维任意域内点集的 Delaunay 三角划分的研究．计算机学报　18（5）：357～364
王新生等　2002．Voronoi 图的扩展、生成及其应用于界定城市空间影响范围　华中师范大学学报（自然科学版），36（1）：107－111
徐寅峰，杨波艇　1995．Voronoi 图在球面上的推广．工程数学学报　12（2）：93～96
张大卫等　1997．平面连通域 Voronoi 图的算法．计算机辅助设计与图形学学报．9（3）：427～435
张有会．1995．线段加权的 Voronoi 图．计算机学报　18（11）：822～829
赵学胜．2002．基于 QTM 的球面 Voronoi 数据模型研究［博士学位论文］．中国矿业大学（北京校区）
Aonuma H., Imai H., Imai K. and Tokuyama　1990　Maxi/min locations of convex objects and related dynamic Voronoi diagrams. Proc. 6th Annual ACM Symposium on Computational Geometry, Berkeley, California, 225～234
Aurenhammer F. and Edelsbrunner H., 1984. An optimal algorithm for constructing the weighted Voronoi diagram in the plane. Pattern Recongnition. 17 (2): 251～257
Borgefors G., 1986. Distance transformations in digital images. Computer vision, Graphics and Image processing, 34: 344－371
Borgefors G., 1994. Applications using distance transformations. In: Aspects of Visual form Processing, Edited by Arcelli, C., Corella, P. and Boga, G., World Scientific, 616pp
Chou J. J., 1995. Voronoi diagrams for planar shapes. IEEE Computer Graphs and Applications. 15 (2): 52～59
Dutton, G., 1991a. Polyhedral hierarchical tessellations: The shape of GIS to come. Geographical Information Systems, 1 (3): 49～55
Dutton, G., 1991b. Zenethial OrthoTriangular Projection. Proceedings of Auto Carto 10. Falls Church, VA: ACSM, 77－95
Fortune S., 1986. A sweepling algorithm for Voronoi diagrams. Proc. 2nd Annual Sym. on Computational Geometry. NY, 313－332
Held M., 1998. Voronoi diagrams and offset curves of curvilinear polygons. Computer-Aided Design. 30 (4): 287～300
Hwang F., 1979. An O (nlogn) algorithm spanning trees, JACM, 26: 177～182
Kirkpatrick D., Efficient computation of continuous skeletons. Proc. 20th Annual IEEE Symposium on Foundations of Computer Science, 18～27
Lee D. and Drysdale R., 1981. Generational of Voronoi diagram in the plane. SIAM J. Comput., 10 (1): 73～87
Leven D. and Sharir M., 1987. Planning a purely translational motion of a convex robot in two-dimensional space using Voronoi diagrams. Discrete & Comput. Geometry, 2: 9～31
Lukatela H., 1987. Hipparchus Geopositioning Model: An Overview. Proceedings of the Eighth International Symposium on Computer-Assisted Cartography, Baltimore, Maryland, 87～96
Lukatela H., 1989. Hipparchus Data Structure: Points, Lines and Regions in Spherical Voronoi Grid. Proceedings of the Ninth International Symposium on Computer-Assisted Cartography, Baltimore, Maryland. 164～170
Okabe A., Boots B. and Sugihara K., 1992. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, John Wiley and Sons, 532pp
Roos T., 1991. Dynamic Voronoi Diagrams. Dissertation zur Erlangung des Naturwissenschaftlichen Doktorgrades, Wurzburg, 63－73
Sharir, 1985. Intersection and closest pair problems for a set of planar discs. SIAM Journal of Computing, 14: 448～468
Shmos M. and Hoey D., 1975. Closet-point problem. Processing of the 16th Annual IEEE Symposium on Foundation of Computer Science, 151～162
Tokuyama T., 1988. Deformation of merged Voronoi diagrams with translations. TRL Research Report TR87-0049, IBM Tokyo Research Laboratory
Watson D. F., 1988. Natural Neighbor Sorting on the N-Dimensional Sphere, Pattern Recognition, 21 (1): 63～67

Watson D. F., 1994. An implementation of natural neighbor interpolation. Published by David Watson, P. O. Box 734, Claremont, WA6010, Australia, 170

Watson D. F, 1998. Modemap: An Implementation of Natural Neighbor Interpolation on the Sphere. http://members.iinet.net.au/ watson/modemap.html

Yang W. P. and Gold C. M., 1996. Managing spatial objects with VMO-tree. In Proc. 7th Int. Sym. on spatial data handling. Delft, Taylor & Francis, 711~726

Yang W. P., 1998. The design of a dynamic Voronoi map object (VMO) model for sustainable forestry data management. Ph. D dissertations of University Laval

Yap C., 1987. An O (nlogn) algorithm for the Voronoi diagram of a set of simple curve segments. Discrete & Computational Geometry, 2: 365~393

第十七章　空间变换算法

空间变换是GIS数据处理、交换与共享的一项重要内容，更是不同数据源集成分析应用的重要基础。空间变换包括三个方面的主要内容，即不同平面坐标系之间的数据变换，不同地理投影之间的数据转换以及可视化过程中不同视场（视点、视角）之间的坐标变换。

§17.1　地图坐标变换算法

各种地图的数字化仍然是GIS数据输入的重要来源。当使用数字化仪对地图进行图数转换时，由于数字化仪桌面坐标系与地图直角坐标系的不一致以及图纸变形等原因，由数字化仪读出的坐标（x，y）不可能与地图直角坐标系中的坐标（X，Y）一致。为了获得一致的数据，必须进行坐标转换以使坐标系统一。坐标变换主要涉及相似变换和仿射变换。

17.1.1　相似变换算法

相似变换是假设地图的实际比例尺在X、Y方向是一致的。如图17-1所示，设XOY为地图直角坐标系，$xo'y$为数字化仪坐标系。两坐标系之间坐标轴的交角为α，O'相对于O的平移距离为A_0，B_0，地图的实际比例尺在X、Y方向均为m。则根据图形学原理，其坐标变换公式为：

$$
\begin{aligned}
X &= m(x\cos\alpha - y\sin\alpha) + A_0 \\
Y &= m(x\sin\alpha + y\cos\alpha) + B_0
\end{aligned}
\tag{17-1}
$$

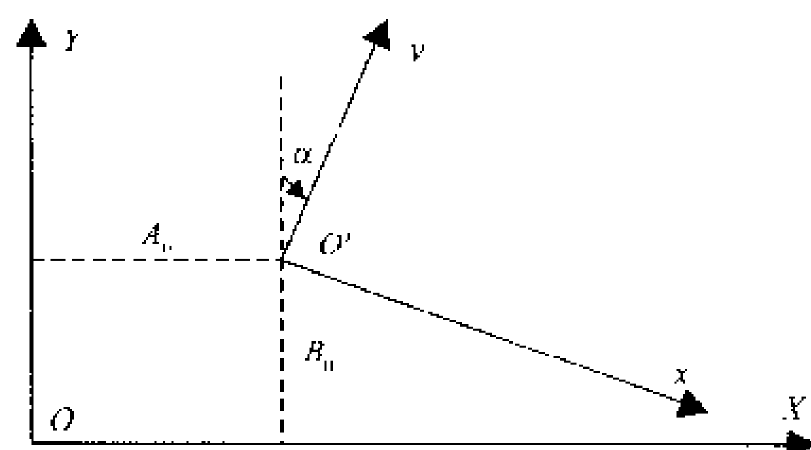


图17-1　坐标相似变换原理

若令$A_1 = m\cos\alpha$，$B_1 = m\sin\alpha$，则上式可简化为

$$
\begin{aligned}
X &= A_0 + A_1x - B_1y \\
Y &= B_0 + B_1x + A_1y
\end{aligned}
\tag{17-2}
$$

设 Q_x、Q_y 分别表示某些已知点（数字化控制点）的已知 X、Y 实际坐标与按上式转换的计算坐标之差，即：

$$
\begin{aligned}
Q_x &= X - (A_0 + A_1 x - B_1 y) \\
Q_y &= Y - (B_0 + B_1 x + A_1 y)
\end{aligned}
\tag{17-3}
$$

根据最小二乘原理，以 $[Q_x^2 + Q_y^2]$ 最小为条件，得到法方程：

$$
\begin{cases}
\sum X = A_0 n + A_1 \sum x - B_1 \sum y \\
\sum Y = B_0 n + A_1 \sum y + B_1 \sum x \\
\sum xX + \sum yY = A_0 \sum x + B_0 \sum y + A_1 \sum (x^2 + y^2) \\
\sum xY - \sum yX = B_0 \sum x - A_0 \sum y + B_1 \sum (x^2 + y^2)
\end{cases}
\tag{17-4}
$$

应用消元法求解以上法方程，可确定各系数如下：

$$
\begin{cases}
A_0 = \bar{X} - A_1 \bar{x} + B_1 \bar{y} \\
B_0 = \bar{Y} - A_1 \bar{y} - B_1 \bar{x} \\
A_1 = \dfrac{\sum (x - \bar{x})(X - \bar{X}) + \sum (y - \bar{y})(Y - \bar{Y})}{\sum (x - \bar{x})^2 + (y - \bar{y})^2} \\
B_1 = \dfrac{\sum (x - \bar{x})(Y - \bar{Y}) + \sum (y - \bar{y})(X - \bar{X})}{\sum (x - \bar{x})^2 + (y - \bar{y})^2}
\end{cases}
\tag{17-5}
$$

式中：n 为已知点总数；$\bar{X}$ 为已知点的实测 X 坐标均值，$\bar{X} = (\sum X)/n$；$\bar{Y}$ 为已知点的实测 Y 坐标均值，$\bar{Y} = (\sum Y)/n$；$\bar{x}$ 为已知点的数字化 x 坐标均值，$\bar{x} = (\sum x)/n$；$\bar{y}$ 为已知点的数字化 y 坐标均值，$\bar{y} = (\sum y)/n$。

17.1.2 仿射变换算法

实际上，由于地图图纸纹理方向性质的差异，图纸日久变形后将导致在 X、Y 方向上的变形不一致，使得图纸数字化时 X、Y 方向的比例尺存在差异。从而需要进行仿射变换。假设变形后地图的实际比例尺在 X、Y 方向分别为 m_x 和 m_y。则根据图形学原理，式（17-1）可写为：

$$
\begin{aligned}
X &= (m_x \cos\alpha) x - (m_x \sin\alpha) y + A_0 \\
Y &= (m_y \sin\alpha) x + (m_y \cos\alpha) y + B_0
\end{aligned}
\tag{17-6}
$$

若令 $A_1 = m_x \cos\alpha$，$A_2 = m_x \sin\alpha$，$B_1 = m_y \sin\alpha$，$B_2 = m_y \cos\alpha$ 则上式可简化为

$$
\begin{aligned}
X &= A_0 + A_1 x - A_2 y \\
Y &= B_0 + B_1 x + B_2 y
\end{aligned}
\tag{17-7}
$$

设 Q_x，Q_y 分别表示某些已知点（数字化控制点）的已知 X、Y 实际坐标与按上式转换的计算坐标之差，即：

$$
\begin{aligned}
Q_x &= X - (A_0 + A_1 x - A_2 y) \\
Q_y &= Y - (B_0 + B_1 x + B_2 y)
\end{aligned}
\tag{17-8}
$$

根据最小二乘原理，以 $[Q_x^2 + Q_y^2]$ 最小为条件，得到两组法方程：

$$
\begin{cases}
\sum X = A_0 n + A_1 \sum x + A_2 \sum y \\
\sum xX = A_0 \sum x + A_1 \sum x^2 + A_2 \sum xy \\
\sum yX = A_0 \sum y + A_1 \sum xy + A_2 \sum y^2
\end{cases}
\tag{17-9}
$$

$$
\begin{cases}
\sum Y = B_0 n + B_1 \sum x + B_2 \sum y \\
\sum xY = B_0 \sum x + B_1 \sum x^2 + B_2 \sum xy \\
\sum yY = B_0 \sum y + B_1 \sum xy + B_2 \sum y^2
\end{cases}
\tag{17-10}
$$

求解以上法方程，可以确定各系数。

§17.2 地图投影变换算法

有时，不同来源的地图还可能存在地图投影与地理坐标的差异，以及地图比例尺之间、地图比例尺与数字化仪的长度单位之间的不一致。为了获得一致的数据，必须进行地图投影变换。投影变换是地图制图的理论基础，是 GIS 区别于其他管理信息系统以及 CAD 的重要方面，主要用来解决换带计算、地图转绘、图层叠加、数据集成等问题。

17.2.1 算法分析

地图投影变换方法可以采用正解变换、反解变换、综合变换、数值变换或数值-解析变换方式。其中前三者都是在已知原投影和新投影解析式条件下，来求得两者之间的解析关系式，一般通称为解析变换；后两者是不知原投影的解析式或不知原投影方程的情况下，采用多项式来确定两者之间的关系。

1. 正解变换

正解变换也称直接变换，是通过建立两种投影之间的严密数学解析关系，直接由一种投影的数字化坐标（x，y）精确变换为另一种投影的坐标（X，Y）。其变换表达式为：

$$
\begin{aligned}
X &= F_1(x, y) \\
Y &= F_2(x, y)
\end{aligned}
\tag{17-11}
$$

2. 反解变换

反解变换是以地理坐标（φ，λ）为中间媒介，先将一种投影坐标（x，y）反解出其地理坐标（φ，λ），然后再将其地理坐标（φ，λ）代入另一种投影的坐标计算公式中，计算出新的投影坐标（X，Y）。

若原投影与地理坐标的解析关系为：$x = f_1(\varphi, \lambda)$，$y = f_2(\varphi, \lambda)$，其反算关系式为：$\varphi = g_1(x, y)$，$\lambda = g_2(x, y)$；新投影与地理坐标的解析式为：$X = F_1(\varphi, \lambda)$，$Y = F_2(\varphi, \lambda)$，则其变换表达式为：

$$X = F_1(g(x,y), g_2(x,y))$$
$$Y = F_2(g(x,y), g_2(x,y)) \tag{17-12}$$

3．综合变换

综合变换是将正解变换与反解变换结合起来使用。通常是先反求出原投影点的地理坐标之一（φ 或 λ），然后根据这一地理坐标与直角坐标之一（x 或 y）相配合求出新投影下点的直角坐标。某些情况下，采用综合变换法要比正解变换和反解变换简便。

4．数值变换

数值变换是根据两种投影在变换区域内的若干同名数字化点，采用多项式逼近的方法来建立两者之间的关系，从而实现两者之间的坐标变换。理论证明，并不是多项式的次数越高，计算精度就越高，往往是多项式的次数过高时拟合结果会出现“抖动”现象，而且多项式次数过高将增加算法的复杂度，并影响计算速度。一般选择多项式的次数为 3 次，其一般表达为：

$$\begin{aligned} X &= a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + a_{30}x^3 + a_{21}x^2y \\ &\quad + a_{12}xy^2 + a_{03}y^3 \\ Y &= b_{00} + b_{10}x + b_{01}y + b_{20}x^2 + b_{11}xy + b_{02}y^2 + b_{30}x^3 \\ &\quad + b_{21}x^2y + b_{12}xy^2 + b_{03}y^3 \end{aligned} \tag{17-13}$$

为解算上述 3 次多项式，需要在两投影之间选择若干同名基准点组成法方程来确定其中的待定系数。理论上说，采用该算法时，除基准点之外的其他点上都或多或少存在误差，离基准点越近的点误差越小，而离基准点越远的点误差越大。因此，基准点的选择应符合以下要求：

(1) 在整个研究范围内分布均匀；

(2) 具有代表性，可以选择突出的标记点或特殊的特征点。

李国建等（2001）针对任意两个投影平面之间的坐标变换问题，研究提出了一种通用的基于双仿射微分纠正的直接变换算法，并指出该法特别适合于 DRG 等栅格地图上快速投影变换。

5．数值-解析变换

数值-解析变换是先采用多项式逼近的方法确定原投影的地理坐标，然后将所确定的地理坐标代入新投影与地理坐标之间的解析式中，求得新投影的坐标，从而实现两种投影之间的变换。多项式逼近形式为：

$$\begin{cases} \varphi = \sum_{i=0}^{n}\sum_{j=0}^{n} a_{ij}x^iy^j \\ \lambda = \sum_{i=0}^{n}\sum_{j=0}^{n} b_{ij}x^iy^j \end{cases} \cdots\cdots(i + j \leqslant n) \tag{17-14}$$

式中：n 为多项式的次数。

上述 5 种地图投影变换算法的适用情况和特点见表 17-1。

表 17-1　投影变换算法的特点与适用范围

<table>
<tr><th colspan="2">算法名称</th><th>主要特点</th><th>适用范围</th></tr>
<tr><td rowspan="3">解析变换</td><td>正解变换</td><td>（1）表达了制图过程的数学实质，不同投影之间具有精确的对应关系；（2）在解决多投影问题时，存在计算冗余问题（若 n 种投影，则需（n-1）种投影变换方式）</td><td>受制图区域影响</td></tr>
<tr><td>反解变换</td><td>方法严密、不受区域大小影响</td><td>任何情况</td></tr>
<tr><td>综合变换</td><td>（1）某些情况下比单一使用正解变换或反解变换更简便；（2）并不是所有的投影均能通过地理坐标与直角坐标的配合来进行投影变换</td><td>视情况而定</td></tr>
<tr><td colspan="2">数值变换</td><td>（1）不能反映投影的数学实质；（2）不能进行全区域的投影变换，常采用分块处理的办法，则给计算机自动处理（如何分块、块的大小）带来困难</td><td>局部区域</td></tr>
<tr><td colspan="2">数值-解析变换</td><td>同上</td><td>同上</td></tr>
</table>

17.2.2　算法实现过程

进行地图投影变换时，不同的投影方式需要不同的参数，如割圆锥投影的变换必须指明两条标准纬线，高斯投影时需要确定中央经线等。李英奎等（2000）将不同投影所采用的参数进行综合，统一定义为投影参数并保持统一的形式，如表 17-2 所示。

表 17-2　主要投影参数

（李英奎等，2000）

序号	参数名称	代表符号	备注说明
1	投影名称	PrjName	用于用户界面的投影描述
2	投影类型	PrjType	系统内部惟一确定的投影标识
3	投影椭球体	Ellipse	椭球体名称，用来确定椭球体参数
4	投影单位	Unit	投影变换的坐标单位
5	原点经度	OriginLongiture	投影的坐标原点经度，高斯投影中代表中央经线
6	原点纬度	OriginLatitude	投影的坐标原点纬度
7	第一标准纬线	StdLatitude1	用于圆锥投影
8	第二标准纬线	StdLatitude12	用于圆锥投影

投影变换的计算机实现过程包括初始化、投影变换和后处理 3 个主要步骤，如图 17-2 所示。

由于正解变换、综合变换、数值变换和数值-解析变换均存在不同程度的缺陷，而反解变换不仅能反映投影的数学实质，而且不受地理区域的限制，因而实际应用中多选择反解变换。对于 n 种地图投影之间的变换问题，只需在程序中设置 n 种由直角坐标到

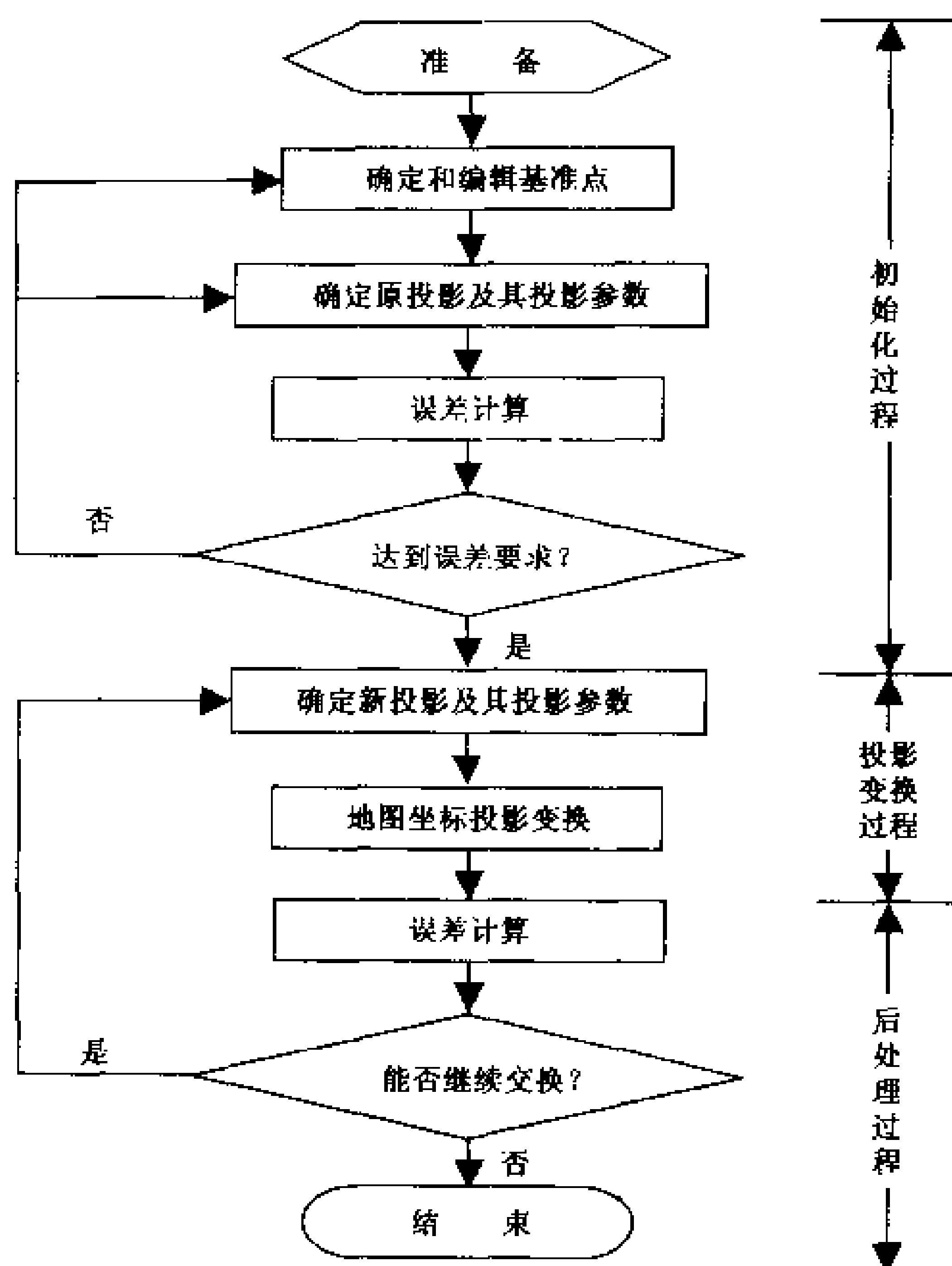


图 17-2 地图投影变换过程
（李英奎等，2000）

地理坐标和n种由地理坐标到直角坐标的变换公式即可；以后每增加一种投影方式，只需增加一对投影变换公式即可。

李英奎等（2000）在多种投影之间利用反解变换对地图坐标的逐点变换进行了过程优化和程序开发，提高了变换效率。其算法过程如图 17-3 所示。

§17.3 透视投影变换算法

平行投影是我们所熟悉的一种投影方式，在机械制图、工程制图、美术设计等领域广泛应用。由于GIS研究的是大范围的地理区域，透视投影更符合人的视觉习惯，因而GIS的可视化与地形模拟等过程一般采用透视投影。透视投影按照主灭点的个数分为一点透视、两点透视和三点透视，一点透视和两点透视如图 17-4 所示。所谓灭点，即任何一束不平行于投影平面的平行线的透视投影的汇聚点；当灭点位于坐标轴上时，称之为主灭点。主灭点数是和投影平面切割坐标轴的数量相对应的，若投影平面仅切割z轴，则z轴是投影平面的法线，在z轴上有一个主灭点；而平行于x轴或y轴的直

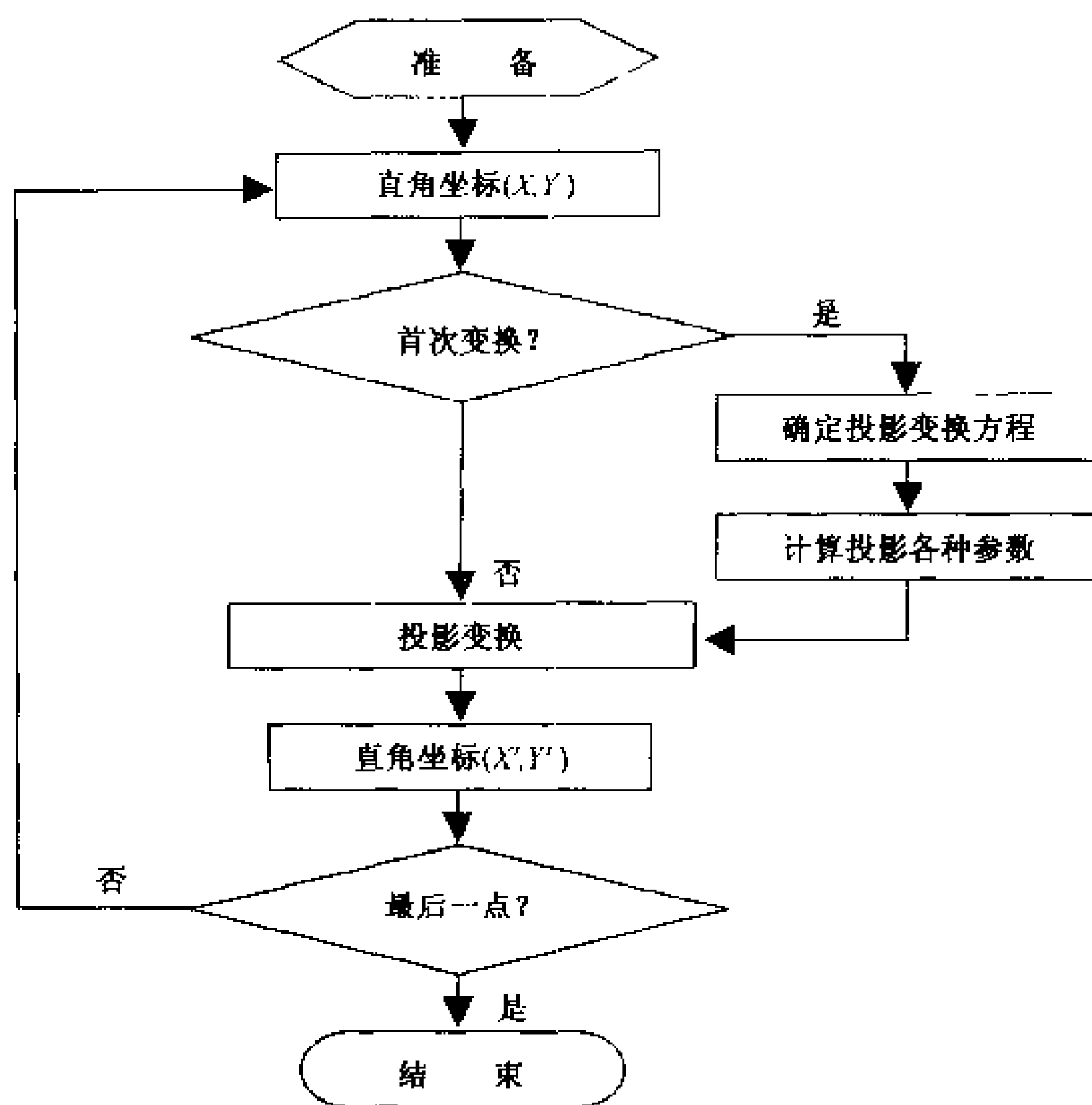


图 17-3 基于反解变换的地图投影变换优化过程
（李英奎等，2000）

线也平行于投影面，但没有灭点。

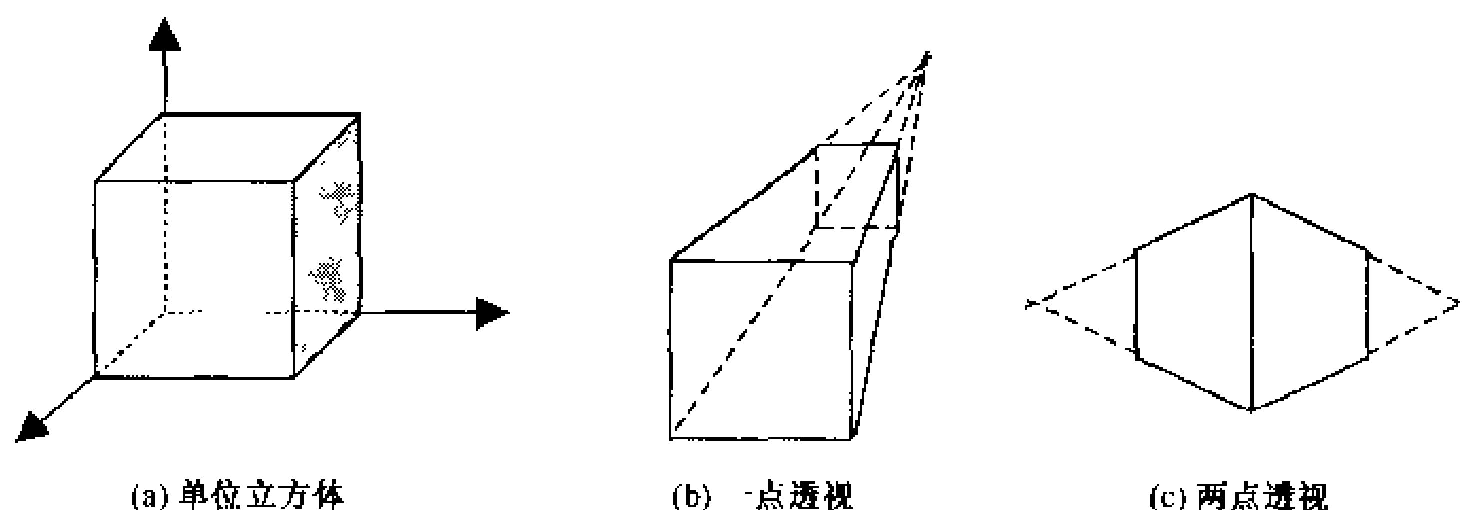


图 17-4 单位立方体的一点透视和两点透视
（孙家广，1998）

17.3.1 简单的一点透视变换算法

首先介绍简单的一点投影，即投影面取为坐标系的一个平面。如图 17-5 所示，设透视投影的视点为 $P_c(x_c, y_c, z_c)$，投影平面为 XOY 平面，形体上一点 $P(x, y, z)$的投影为 $P_s(x_s, y_s)$。现推导求 $P_s(x_s, y_s)$的公式如下（孙家广，1998）。

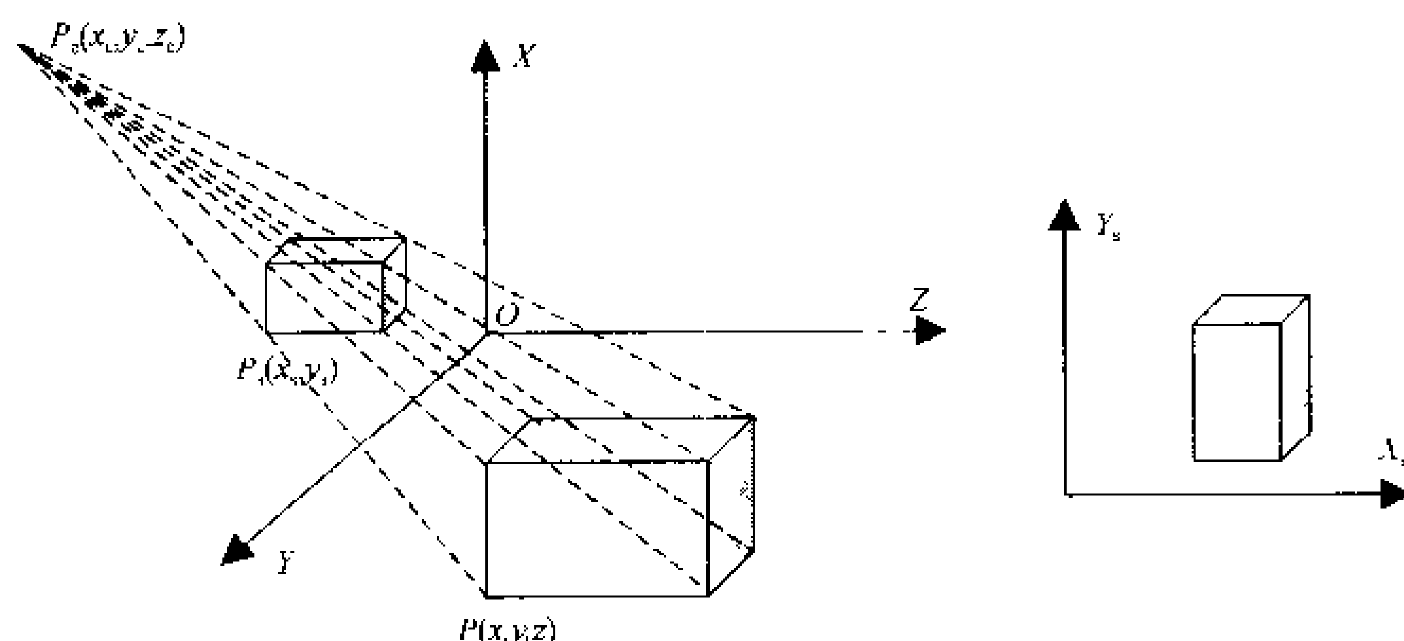


图 17-5 简单的一点透视
（孙家广，1998）

由两点式可得投影线 P_cP 的方程：

$$\begin{cases} x_s = x_c + (x - x_c)t \\ y_s = y_c + (y - y_c)t \\ z_s = z_c + (z - z_c)t \end{cases} \tag{17-15a}$$

由 $z_s=0$ 可知：$t = -\frac{z_c}{z - z_c}$，代入上式得：

$$\begin{cases} x_s = (x_c z - x z_c)/(z - z_c) \\ y_s = (y_c z - y z_c)/(z - z_c) \end{cases} \tag{17-15b}$$

17.3.2 复杂的一点透视变换算法

在上述简单的一点透视投影变换中，虽然得出的投影变换公式直观、好理解，但实际操作困难，因为用户要选择一个较好的视点来满足简单的一点透视投影条件并不容易。而通常，人们习惯于以物体为中心（即物体不动），环绕物体从各个不同的方向来进行观察，这就是复杂的一点透视投影变换问题，或称任意的一点透视变换。

为了推导投影变换公式，需要引入一个过渡坐标系，称为观察坐标系 $X_eY_eZ_e$，如图 17-6 所示。然后，将用户坐标系 $X_wY_wZ_w$ 下的坐标点（x_w，y_w，z_w）变换为观测坐标系 $X_eY_eZ_e$ 下的坐标（x_e，y_e，z_e）。再利用简单一点透视投影公式，即可求得形体上一点 P（x_w，y_w，z_w）在参考坐标系中的坐标（x_e，y_e，z_e），变换公式为：

$$[x_e \quad y_e \quad z_e \quad 1] = [x_w \quad y_w \quad z_w \quad 1] \cdot V \tag{17-16}$$

关键是确定变换矩阵 V，如图 17-7 所示，V 的推导经以下 6 步（孙家广，1998）：

(1) 将用户坐标系的原点平移到视点（设视点在用户坐标系下的坐标为（a，b，c）)，得平移变换矩阵 T_1：

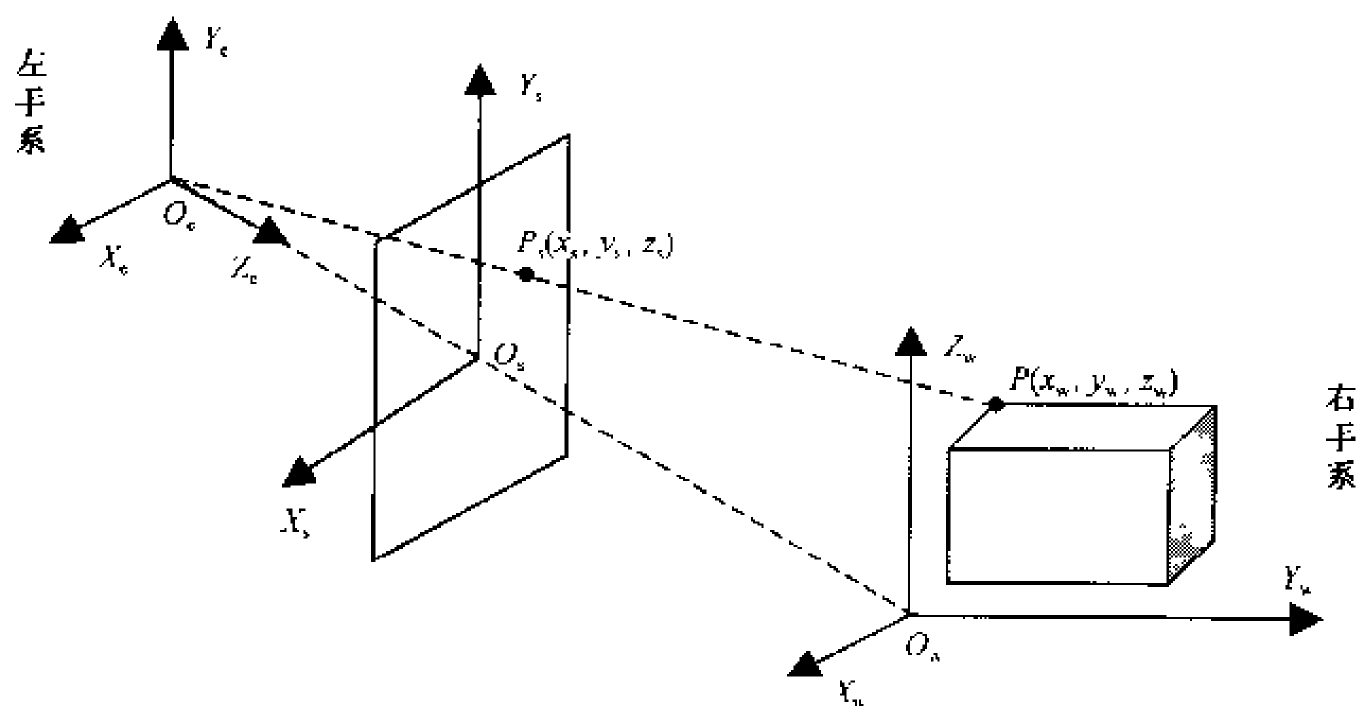


图 17-6 复杂的一点透视

（孙家广，1998）

$$
T_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & -b & c & 1 \end{bmatrix} \tag{17-17a}
$$

（2）令平移后的新坐标系绕其 X' 轴旋转 90°，即物体上的点顺时针旋转 90°，得旋转变换矩阵 T_2：

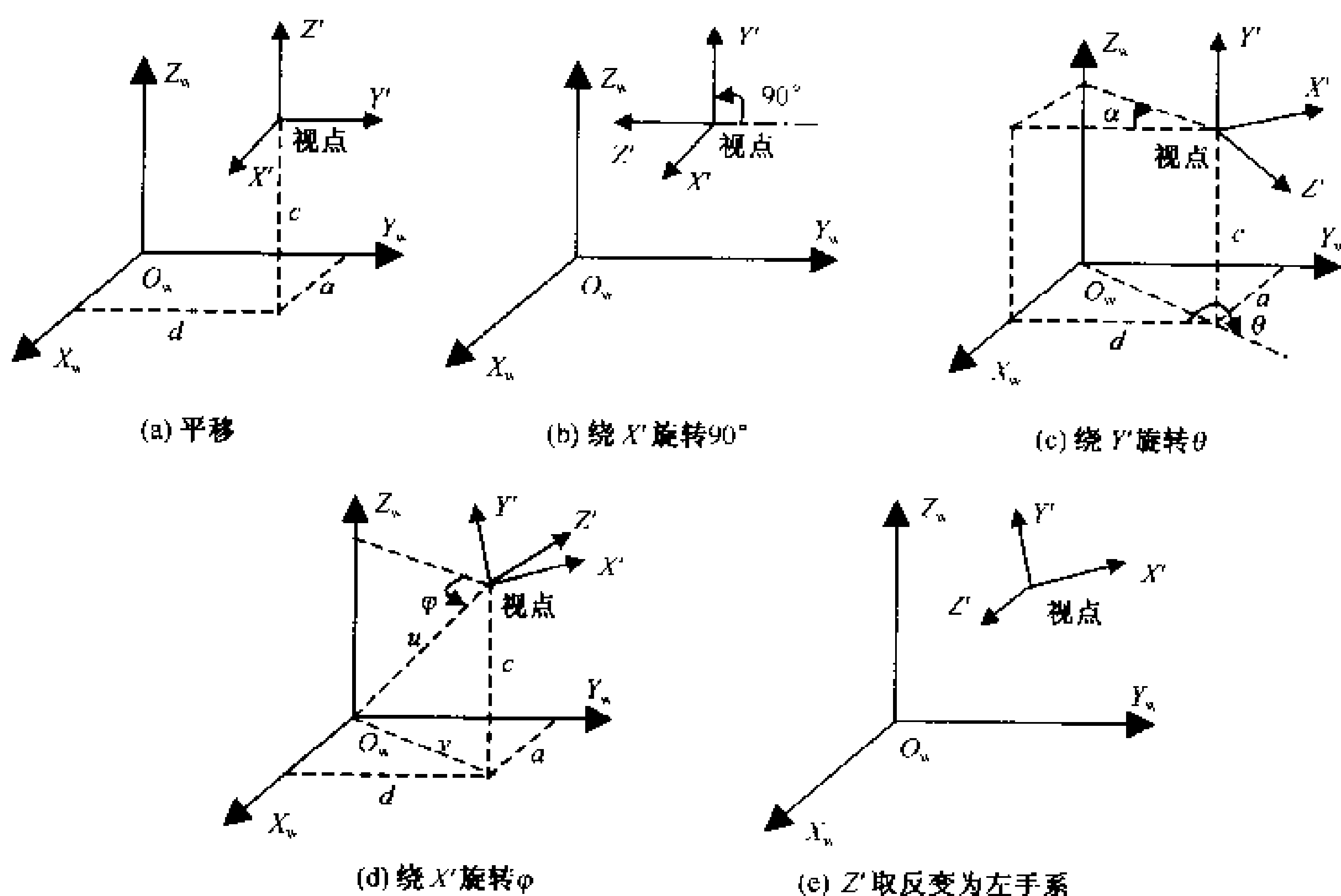


图 17-7 从用户坐标到观察坐标的变换过程

（孙家广，1998）

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 90^\circ & -\sin 90^\circ & 0 \\ 0 & \sin 90^\circ & \cos 90^\circ & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17-17b}$$

(3) 再将新坐标系绕其 Y' 轴顺时针转 θ 角（此时 θ 角大于 180°），即物体上的点逆时针旋转 θ 角，此时 $\cos\theta = \dfrac{b}{\sqrt{a^2+b^2}}$，$\sin\theta = \dfrac{a}{\sqrt{a^2+b^2}}$，令 $v = \sqrt{a^2+b^2}$，得旋转变换矩阵 T_3：

$$T_3 = \begin{bmatrix} -b/v & 0 & a/v & 0 \\ 0 & 1 & 0 & 0 \\ -a/v & 0 & -b/v & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17-17c}$$

(4) 再将新坐标系绕其 x' 轴顺时针转 φ 角，即物体上的点逆时针旋转 φ 角，此时 $\cos\varphi = v/u$，$\sin\varphi = c/u$，令 $u = \sqrt{a^2+b^2+c^2}$，得旋转变换矩阵 T_4：

$$T_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & v/u & c/u & 0 \\ 0 & -c/u & v/u & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17-17d}$$

(5) 再将右手系变成左手系，即令 z' 轴反向，得变换矩阵 T_5：

$$T_5 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17-17e}$$

(6) 将以上各变换矩阵相乘，即得变换矩阵 V：

$$V = T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot T_5 = \begin{bmatrix} -b/v & ac/uv & -a/u & 0 \\ a/v & -bc/uv & b/u & 0 \\ 0 & v/u & -c/u & 0 \\ 0 & 0 & u & 1 \end{bmatrix} \tag{17-17f}$$

设观测坐标系原点 O_e（即视点）离视平面原点 O_s 和用户坐标系原点 O_w 的距离分别为 d_s 和 d_w，则根据比例关系即可在已求得的（x_e，y_e，z_e）的基础上，按下式计算得到物体上点 P（x_w，y_w，z_w）在视平面上的投影坐标（x_s，y_s）：

$$\begin{cases} x_s = x_e \cdot \dfrac{d_s}{d_w} \\ y_s = y_e \cdot \dfrac{d_s}{d_w} \end{cases} \tag{17-17g}$$

孙家广等还指出，按上述变换公式设计算法程序时，还应注意两个问题：

(1) 为了确保视点任意移动时均可以看到物体不同位置的形状，可以先求出物体的外接球，并将用户坐标系的原点先行平移到外接球的球心；

(2) 在屏幕上显示视平面上的投影图时，仍需作窗口视图变换。

参 考 文 献

黄杏元．1989．地理信息系统概论．北京：高等教育出版社．47～50

龚健雅．2001．地理信息系统基础．北京：科学出版社．156～158

杨启和．1989．地图投影变换的原理．北京：解放军出版社

胡毓钜，龚建文，黄 伟．1997．地图投影．北京：测绘出版社

李英奎，吕肖庆，李 敬．2000．多投影间地图投影变换实现的途径与优化．地理学与国土研究，16（2）：79～84

李国建，胡 鹏．2001．通用的地图投影数值变换．地球信息科学，（4）：61～65

孙家广．1998．计算机图形学．北京：清华大学出版社．380～384

第十八章　空间度量算法

空间度量是 GIS 的一项基本内容和主要功能，也是 GIS 的常用工具。空间度量的主要内容包括空间距离量算、空间方向计算、面积计算、体积计算和坡度、坡向计算等。

§18.1　空间距离与方向度量算法

18.1.1　基于矢量的距离与方向度量算法

1. 平面度量算法

1）任意两点之间的距离与方向度量

如图 18-1 所示，设有两平面点 $P_i(x_i, y_i)$、$P_j(x_j, y_j)$，其平面距离 D_{ij} 和平面方向 α_{ij}（定义为连线 P_iP_j 与 Y 轴正向的夹角）分别为

$$D_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{18-1}$$

$$\alpha_{ij} = \begin{cases} 0 \cdots\cdots\cdots\cdots \Delta y \geqslant 0, \Delta x = 0 \\ \dfrac{\pi}{2} - \arctan\left|\dfrac{\Delta y}{\Delta x}\right| \cdots\cdots \Delta y \geqslant 0, \Delta x > 0 \\ \dfrac{\pi}{2} + \arctan\left|\dfrac{\Delta y}{\Delta x}\right| \cdots\cdots \Delta y < 0, \Delta x > 0 \\ \pi \cdots\cdots\cdots\cdots \Delta y < 0, \Delta x = 0 \\ \dfrac{3\pi}{2} - \arctan\left|\dfrac{\Delta y}{\Delta x}\right| \cdots\cdots \Delta y < 0, \Delta x < 0 \\ \dfrac{3\pi}{2} + \arctan\left|\dfrac{\Delta y}{\Delta x}\right| \cdots\cdots \Delta y \geqslant 0, \Delta x < 0 \\ (\Delta y = y_j - y_i, x = x_j - x_i) \end{cases} \tag{18-2}$$

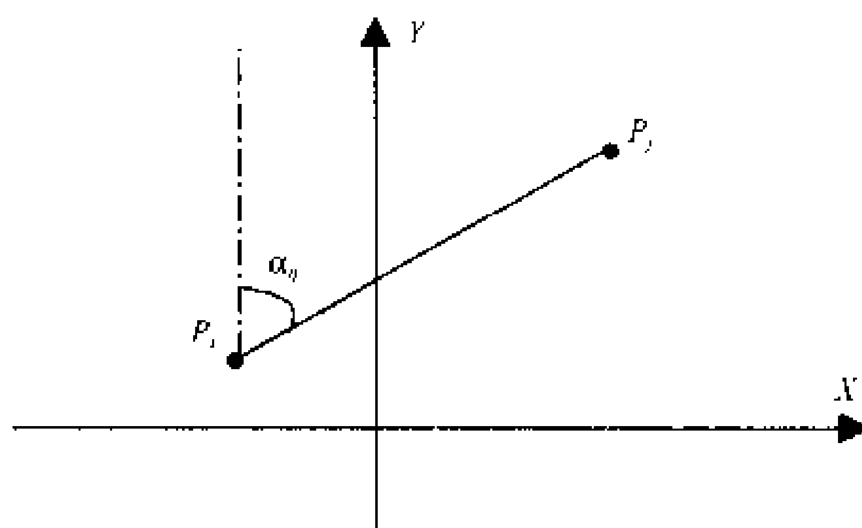


图 18-1　平面距离与方向度量

2) 平面曲线的长度度量

设组成平面曲线的有序点列为 P_1（x_1，y_1）、P_2（x_2，y_2）…P_n（x_n，y_n），该曲线的总长度为 L_{ij}

$$L_{ij} = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{18-3}$$

2. 空间度量算法

1) 任意两点之间的距离与方向度量

如图 18-2 所示，设有两空间点 P_i（x_i，y_i，z_i）、P_j（x_j，y_j，z_j），其空间距离 D_{ij}及其在 XY 平面、XZ 平面和 YZ 平面的投影长度 D_{ij}^{xy}、D_{ij}^{xz}、D_{ij}^{yz}分别为

$$D_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \tag{18-4}$$

$$D_{ij}^{xy} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{18-5}$$

$$D_{ij}^{xz} = \sqrt{(x_j - x_i)^2 + (z_j - z_i)^2} \tag{18-6}$$

$$D_{ij}^{yz} = \sqrt{(y_j - y_i)^2 + (z_j - z_i)^2} \tag{18-7}$$

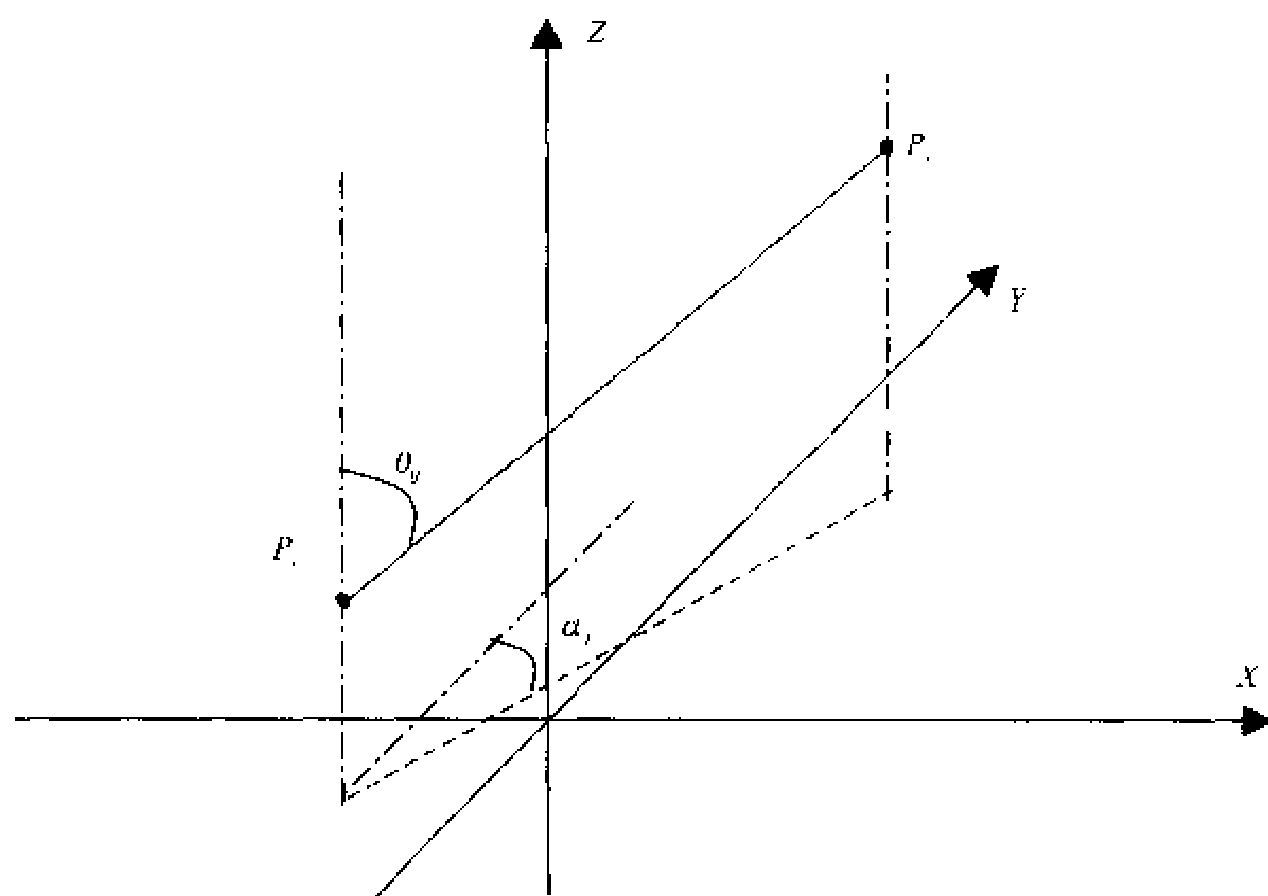


图 18-2　空间距离与方向度量

空间任意两点 P_i（x_i，y_i，z_i）、P_j（x_j，y_j，z_j）之间的方向可以用两个角度参数来表示（θ_{ij}，α_{ij}），其中 θ_{ij}为连线P_iP_j与 Z 轴正向的夹角，α_{ij}为连线P_iP_j在 XY 平面的投影线与 Y 轴正向的夹角，α_{ij}的算法公式同式（18-2），θ_{ij}的算法公式为：

$$\alpha_{ij}=\begin{cases}0, & 当\ \Delta z>0, D_{ij}^{xy}=0\\ \dfrac{\pi}{2}-\arctan\dfrac{\Delta z}{D_{ij}^{xy}}, & 当\ \Delta z\geqslant 0, D_{ij}^{xy}\neq 0\\ \dfrac{\pi}{2}+\arctan\dfrac{|\Delta z|}{D_{ij}^{xy}}, & 当\ \Delta z<0, D_{ij}^{xy}\neq 0\\ \pi, & 当\ \Delta z<0, D_{ij}^{xy}=0\\ (\Delta z=z_j-z_i, D_{ij}^{xy}\ 按式(18\text{-}5)计算) & \end{cases} \tag{18-8}$$

2）空间曲线的长度度量

设组成空间曲线的有序点列为 $P_1(x_1,y_1,z_1)$、$P_2(x_2,y_2,z_2)$、…、$P_n(x_n,y_n,z_n)$，该曲线的总长度为 L_{ij}

$$L_{ij}=\sum_{i=1}^{n-1}\sqrt{(y_{i+1}-y_i)^2+(x_{i+1}-x_i)^2+(z_{i+1}-z_i)^2} \tag{18-9}$$

3．球面度量算法

1）任意两点之间的距离度量

在地球椭球体上，两地理点之间的实地距离不能简单地以平面投影后的欧氏距离来计算。通常，人们以大圆弧长来定义地球表面上两点之间的球面距离。所谓大圆，即经过球心的平面与地球表面相交得到的圆。所谓大圆弧长，即为经过两地理点的大圆上的两点之间的较短弧的长度。

如图 18-3 所示，设两地理点的地理经纬坐标分别为 P_1（φ_1，λ_1）、P_2（φ_2，λ_2），则根据球面三角余弦定理可推知，球面弧段 P_1P_2 的长度为：

$$P_1P_2=R\cdot\arccos P_1P_2 \tag{18-10}$$

式中，R 为地球半径；$\cos P_1P_2$ 为球面弧段 P_1P_2 的余弦值：

$$\cos P_1P_2=\cos(90°-\varphi_1)\cdot\cos(90°-\varphi_2)+\sin(90°-\varphi_1)\cdot\sin(90°-\varphi_2)\cdot\cos(\lambda_1-\lambda_2)$$

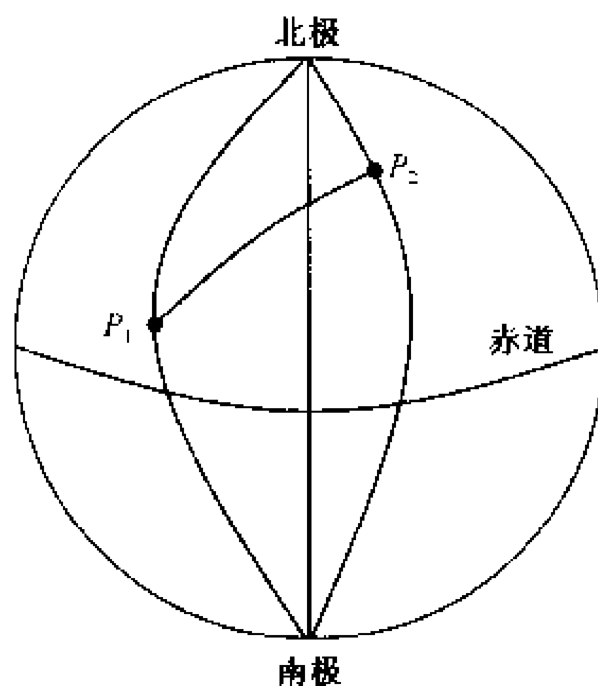


图 18-3　球面距离度量

2）任意两点之间的方向度量

根据大地测量学的有关公式，如图 18-3 所示，P_2（φ_1，λ_1）相对于 P_1（φ_1，λ_1）、P_1（φ_1，λ_1）相对于 P_2（φ_1，λ_1）的方位角 $\alpha_{P_1P_2}$、$\alpha_{P_2P_1}$ 分别为：

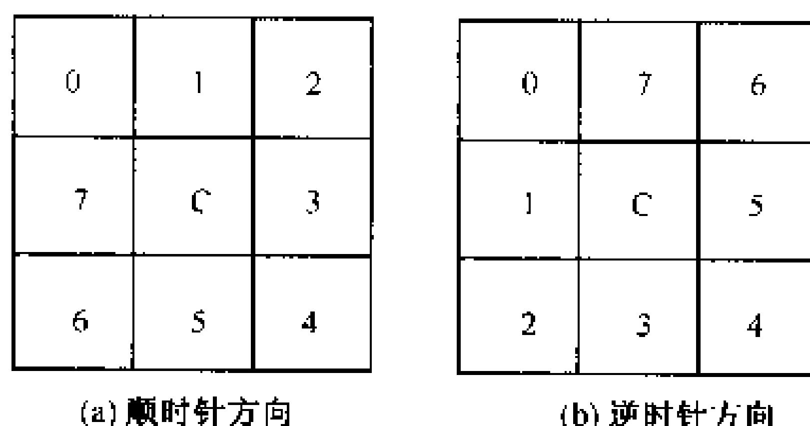
$$\begin{cases}\alpha_{P_1P_2}=\operatorname{arccot}\dfrac{\sin\varphi_2\cdot\cos\varphi_1-\cos\varphi_2\cdot\sin\varphi_1\cdot\cos(\lambda_2-\lambda_1)}{\cos\varphi_2\cdot\sin(\lambda_2-\lambda_1)}\\ \alpha_{P_2P_1}=\operatorname{arccot}\dfrac{\sin\varphi_1\cdot\cos\varphi_2-\cos\varphi_1\cdot\sin\varphi_2\cdot\cos(\lambda_1-\lambda_2)}{\cos\varphi_1\cdot\sin(\lambda_1-\lambda_2)}\end{cases} \tag{18-11}$$

18.1.2 基于栅格的距离与方向度量算法

可以将栅格结构的行列方向分别看成矢量结构的 X 轴和 Y 轴，则栅格结构时点的行列坐标可以等同于矢量结构的 x、y 坐标。因而，前述公式（18-1）～式（18-3）可以无条件地适用。

1. 基于链码的距离度量算法

链码是栅格数据结构中一种基于 3×3 窗口的虚拟编码方式，即以动态的当前栅格为中心，固定从某一位置开始按顺时针或逆时针方向依次对其 8 邻域编码为 0，1，2，3，4，5，6 和 7，如图 18-4 所示。



(a) 顺时针方向　　(b) 逆时针方向

图 18-4　链码编码方式

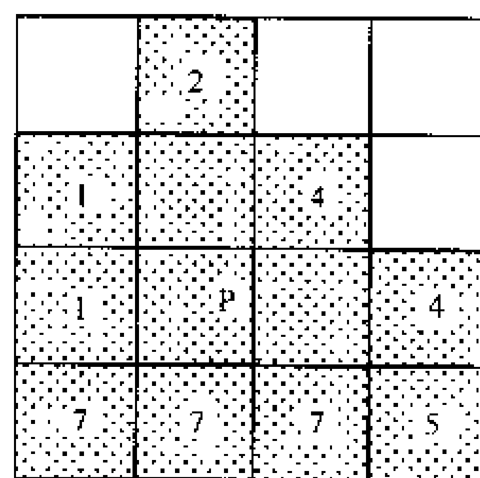


图 18-5　面域的顺时针链码

以如图 18-5 所示的面域多边形边界为例，基于顺时针链码的多边形周长计算过程为：

（1）从边界的某一点（如 0 行 1 列；注：行列编号从 0 开始）开始，将其作为当前 3×3 窗口的中心，查找边界前进方向下一栅格的链码并记录之；

（2）以该栅格为当前 3×3 窗口的中心，继续查找边界前进方向下一栅格的链码并记录之；直到回到起点，并完成链码记录。

（3）打开链码记录并顺序取出，当链码为偶数时，取线段长为$\sqrt{2}d$；当链码为奇数时，取线段长为 d。其中 d 为栅格的尺寸。将以上线段值累加，即为多边形周长。

基于链码的距离与多边形周长计算的统一公式为：

$$L = (m + n\sqrt{2})d \tag{18-12}$$

式中，m 为链码序列 TC 中的奇数总量；n 为链码序列 TC 中的偶数总量；d 为栅格的尺寸。

如图 18-5 所示，沿面域多边形 P 的边界的链码序列 TC 和周长 L 分别为：

$$TC = \{4,4,5,7,7,7,1,1,2\}$$

$$L = (6 + 3\sqrt{2})d$$

2. 方向度量算法

如图 18-6 所示，如果已知所定义的栅格矩阵的纵轴方向与真实地理北方向的交角

θ，则可将按前述公式求得的方向值 α_{ij} 转化为真实地理坐标下的方向值 β_{ij}。转换公式为：

$$\beta_{ij} = \alpha_{ij} + \theta \tag{18-13}$$

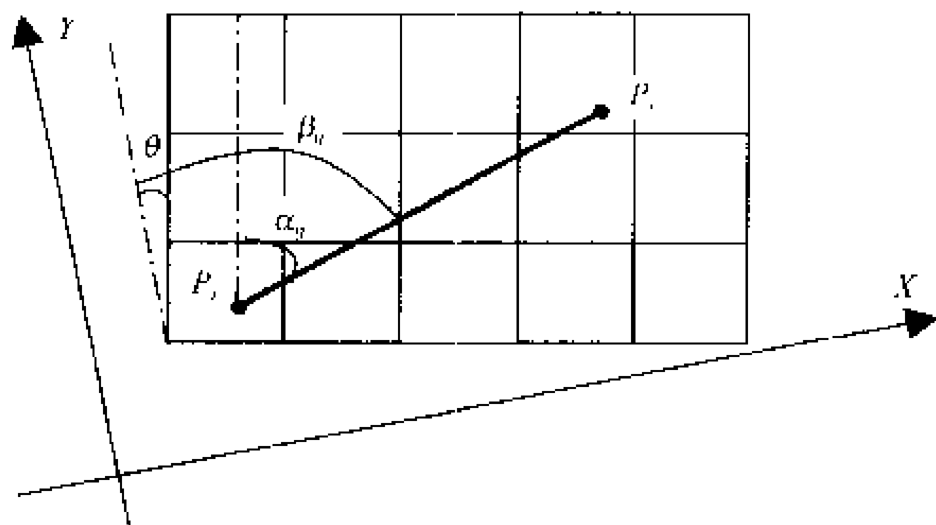


图 18-6　栅格坐标系与地理坐标系中方向度量的对照

§18.2　面积度量算法

18.2.1　基于矢量的面积度量算法

二维矢量环境下，面目标的面积计算有多种方法，常用的是基于积分原理的面积计算。如图 18-7 所示，设面目标的边界为凸多边形，且边界点 P_1、P_2、…、P_n 为顺时针方向排列，则按下式计算的面积为正；若边界点为逆时针方向排列，则按下式计算的面积为负：

$$S = \frac{1}{2}\sum_{i=1}^{n}[(x_{i+1} - x_i)(y_{i+1} + y_i)] \tag{18-14}$$

如果面域边界为非凸多边形，则可以先将其分割为两个或两个以上的凸多边形，分别计算其面积，然后叠加即可，如图 18-8 所示。

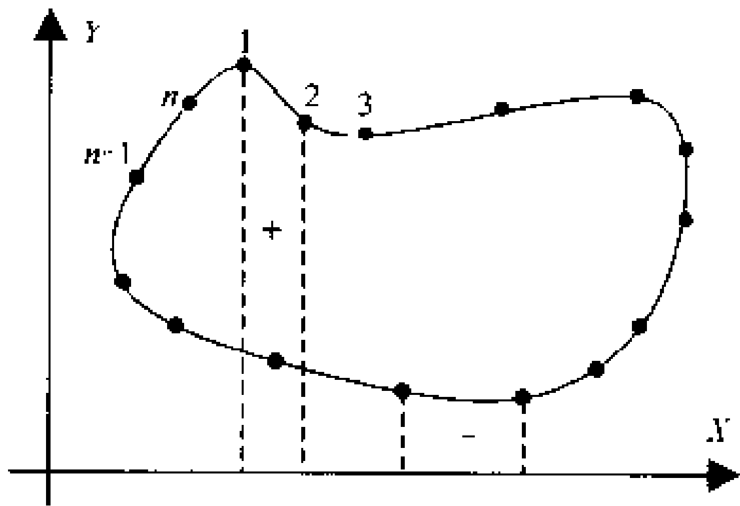


图 18-7　凸多边形面积积分算法原理

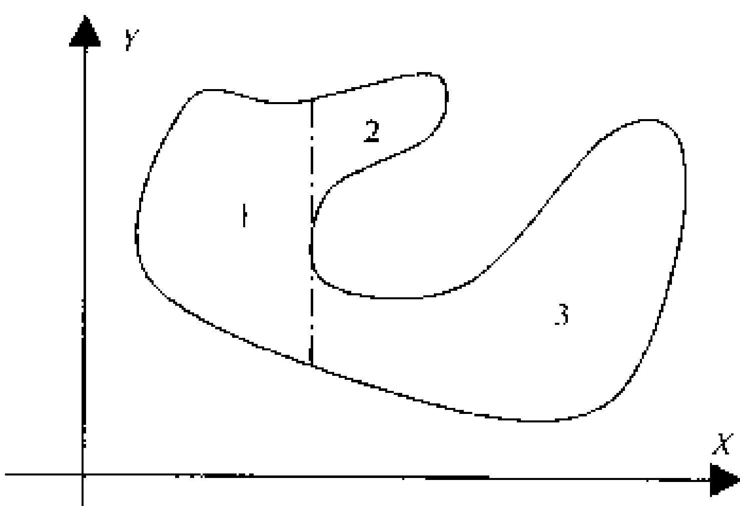


图 18-8　非凸多边形的分割

再介绍一种基于扫描原理的凸多边形面积算法。

如图 18-9 所示，该算法的原理为：设面目标的边界为凸多边形，且边界点顺序排

列为 P_1、P_2、…、P_n、选择可与其余各点通视的某点为扫描基点，从该点向其后继点扫描；每两条相邻扫描线与该两个后继点的连线构成三角形，计算该三角形的面积；依次扫描，直至回到扫描点的前继点。所有扫描三角形的面积之和即为所求的面目标的面积。

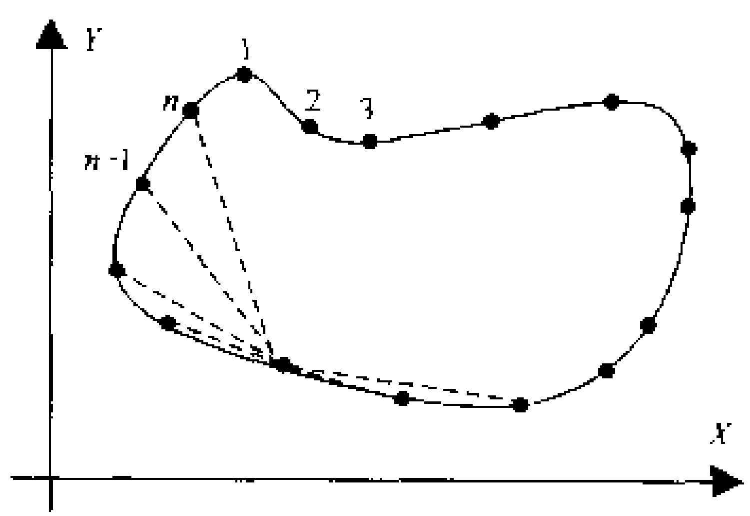


图 18-9 凸多边形面积扫描算法原理

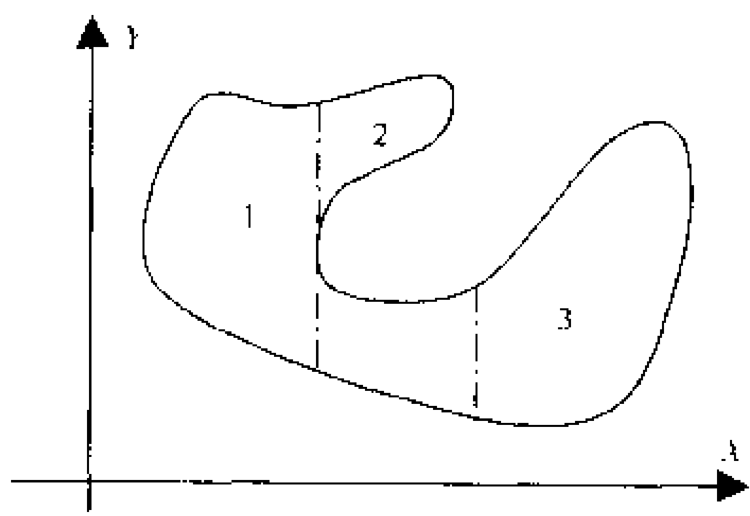


图 18-10 非凸多边形的分割

如果面域边界为非凸多边形，则可以先将其分割为两个或两个以上的凸多边形，分割的前提是保证每个凸多边形均有一个边界点可与其余各边界点通视。分别计算各凸多边形的面积，然后叠加即可，如图 18-10 所示。

由于三维矢量 GIS 尚有许多理论与技术问题有待解决，目前还没有成熟的系统运行。因此，关于三维矢量环境下的空间多边形的表面积度量问题暂不作讨论。基于栅格的三维环境下的空间多边形的表面积度量将在下节中讨论。

18.2.2 基于栅格的面积度量算法

基于栅格的空间曲面的面积度量可以归结为 TIN 表面面积计算和格网 DEM 表面面积计算。而且，格网 DEM 表面积计算可以有两种模式，其一是基于规则的栅格单元的面积累积，其二是将每个格网分解为两个三角形，进而转化为 TIN 进行计算。本节重点介绍基于栅格的平面面积算法、基于 TIN 的区域地形表面积算法、基于 TIN 的区域地形投影面积和基于格网的地形剖面面积 4 个方面。

1. 基于栅格的平面面积算法

基于栅格的平面面积计算有多种不同算法，如基于栅格单元的累积法、基于积分原理的条柱法。基于栅格单元的累积法是在栅格数据记录与属性匹配的基础上，将具有相同属性的同一面域内的栅格单元数进行累计，然后乘以栅格单元面积即可，算法如下：

$$S = N \times S_c \tag{18-15}$$

式中，S 为多边形面积；N 为多边形中栅格总数；S_c 为栅格单元面积。

由于栅格数据往往采用某种压缩编码方式存储，因此，N 的统计要视具体的压缩编码方式而定。如对于游程编码，N 为各游程长度 Length_i 的和；对于四叉树编码，N 为各叶结点大小 Node_i 的和；对于 Morton 压缩编码，N 为各压缩编码段长度 Length_i 之和。

基于积分原理的条柱法是以栅格行（或列）为参考方向，如图 18-11 所示，统计当列号（或行号）相同时，其最大行号与最小行号（或最大列号与最小列号）之差，将所有差数累加并加上差数总数，再乘以栅格单元面积，则得到多边形面积。其算法如下：

$$S = S_c \sum_{i=1}^{n} (R_{i\max} - R_{i\min} + 1) \tag{18-16}$$

式中，$R_{\max}$为对应某一列号的最大行号；$R_{\min}$为对应某一列号的最小行号；n 为条柱总数。

以上是针对多边形区域是凸多边形和无岛多边形的情况，其中 $R_{\max}$和 $R_{\min}$实质上是多边形的上下边界点的行号。当多边形区域为非凸或（和）有岛的复杂多边形区域时，如图 18-12 所示，对应某一列号，可能有多个边界点，此时应将边界点的行号从小到大顺序排列，并不重复地两两组合，再求其差和。算法如下：

$$S = S_c \sum_{i=1}^{n} \sum_{j=1}^{m} (\Delta C_{ij} + 1) \tag{18-17}$$

式中：m 为对应某一列号的边界点组数；ΔC_{ij} 为对应第 i 列的第 j 组边界点的行号之差。

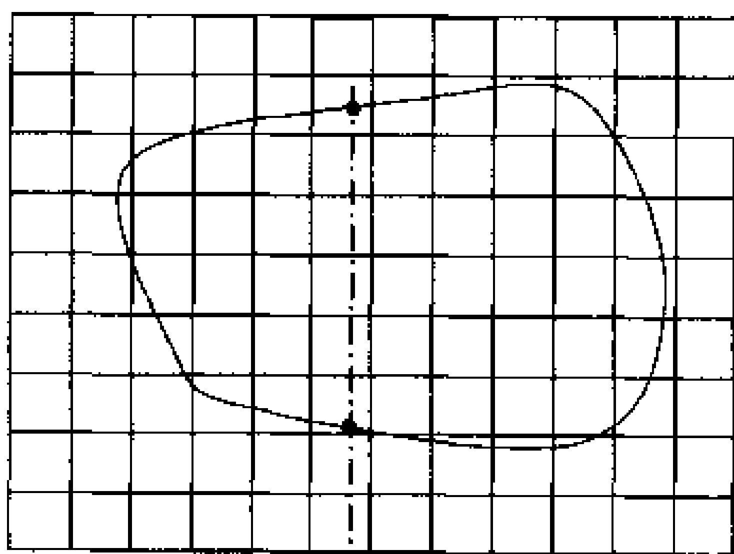


图 18-11 基于积分原理的凸多边形面积算法

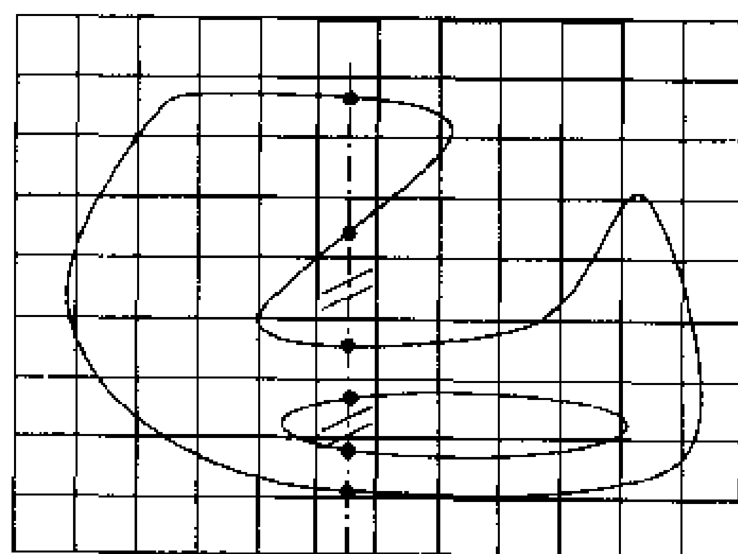


图 18-12 基于积分原理的复杂多边形面积算法

2. 基于 TIN 的表面积算法

地形表面积的计算常基于三角形面积计算原理，即将每一个格网分割为两个三角形。三角形表面积 S 的海伦计算公式为：

$$\begin{cases} S = \sqrt{P(P-D_1)(P-D_2)(P-D_3)} \\ P = \dfrac{1}{2}(D_1 + D_2 + D_3) \\ D_i = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2} \end{cases} \tag{18-18}$$

式中，D_i 为三角形两顶点之间的 3D 空间距离，$i=1\sim3$；P 为三角形周长之半；ΔX、ΔY、ΔZ 为两顶点之间 X、Y、Z 方向的坐标差。

3. 基于 TIN 的投影面积算法

投影面积 S_p 是指任意多边形在水平面上面积，可以直接采用海伦公式计算：

$$\begin{cases} S_p = \sqrt{P(P-D_1)(P-D_2)(P-D_3)} \\ P = \frac{1}{2}(D_1 + D_2 + D_3) \\ D_i = \sqrt{\Delta X^2 + \Delta Y^2} \end{cases} \tag{18-19}$$

式中，D_i 为三角形两顶点之间的平面投影距离，$i=1\sim3$；P 为投影三角形周长之半；ΔX、ΔY 为两顶点之间 X、Y 方向的坐标差。

4. 剖面积算法

剖面积计算是岩土工程、土木工程和地质工程等许多工程领域的一项重要工作。例如，在工程线路设计之后以及在工程实施过程中，需要计算沿线路的剖面面积。设基本参照面（也称基准面）高程为 H_0，则剖面积 S_f 计算公式为：

$$\begin{cases} S_f = \sum_{i=1}^{N-1} \frac{Z_i + Z_{i+1} - 2H_0}{2} \cdot D_{i,i+1} \\ D_{i,i+1} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \end{cases} \tag{18-20}$$

式中，N 为线路与 DEM 格网的交点数（要求按前进方向顺序排列）；Z_i 为第 i 个交点的高程；$D_{i,i+1}$为 P_i、P_{i+1}两个交点之间的平面投影距离。

§18.3 体积度量算法

所谓体积有两种理解，其一为某一空间对象的容积或其所占有空间的体积，其二为空间曲面与某一基准面之间的空间的体积。前者属于立体几何方面的常识，不作讨论；至于后者，随着基准面高程变化，空间曲面的平均高程可能低于基准面，出现负体积的情况。这在工程中称为填方，反之为挖方。山体体积或挖填方体积计算是岩土工程、土木工程和地质工程领域的一项重要工作。通常，可以根据四棱柱、三棱柱体积累计的原理来进行近似计算。其基本思想均是以基底面积（正方形或三角形）乘以格网点曲面的平均高度。然后进行累积，则可求得基于规则格网 DEM 或基于三角形 DEM 的山体体积和挖填方体积。

18.3.1 山体体积算法

如图 18-13 所示，设基本参照面高程为 H_0，则山体体积计算公式分别为：

$$V_3 = \sum_{i=1}^{N} \frac{Z_{i1} + Z_{i2} + Z_{i3} - 3H_0}{3} \cdot S_i \tag{18-21}$$

$$V_4 = \sum_{i=1}^{N} \frac{Z_{i1} + Z_{i2} + Z_{i3} + Z_{i4} - 4H_0}{4} \cdot S_i \tag{18-22}$$

式中，V_3、V_4 分别为基于 TIN 的 DEM 和基于规则格网 DEM 的体积；N 为 DEM 中三角形或规则格网中格网的总数；Z_{ij} 为第 i 个 TIN 或规则格网的角点的高程（$i=1\sim N$）；对于三角形：$j=1\sim3$；对于规则格网：$j=1\sim4$；S_i 为第 i 个三角形或规则格

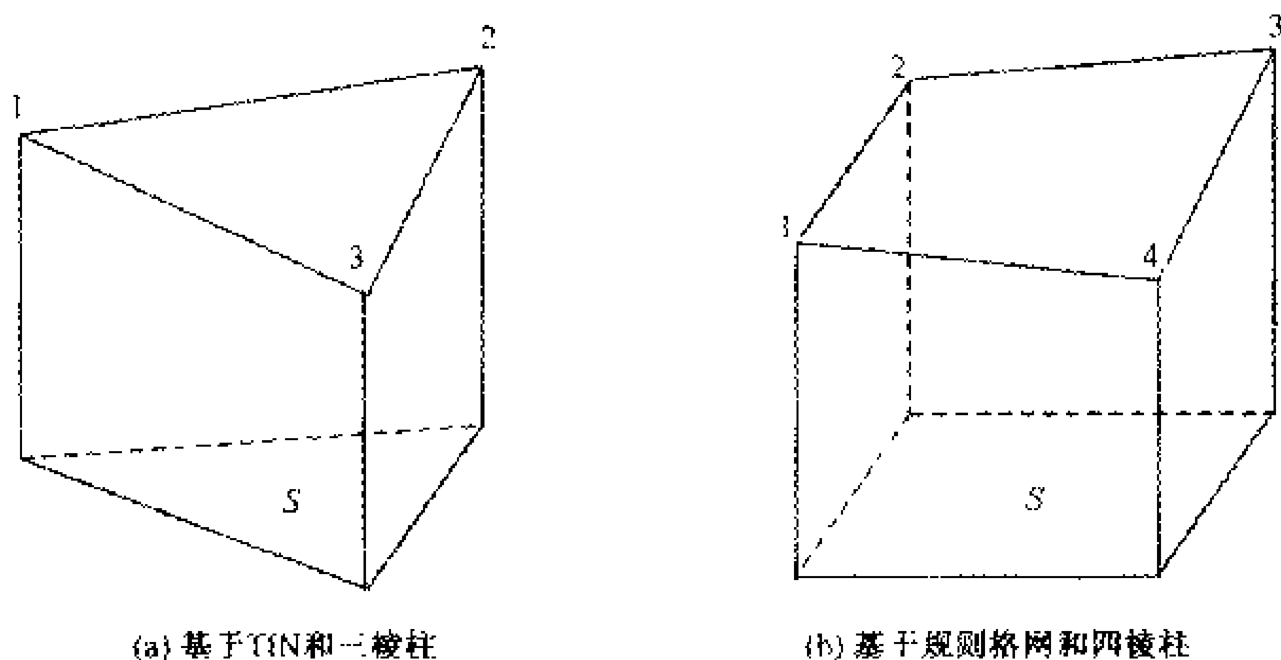


图 18-13　山体体积计算原理

网的投影面积（$i=1\sim N$）。

18.3.2　挖填方体积算法

若已知挖填前后的山体体积分别为 V_0 和 V_a，则挖填方体积为

$$V = V_0 - V_a \tag{18-23}$$

当 $V>0$ 时，表示挖方；当 $V<0$ 时，表示填方。当 $V=0$ 时，表示挖填相当。

§18.4　坡度坡向度量算法

坡度、坡向是地形描述中的常用参数，在各类工程活动和土地开发利用过程需要经常使用。比如，一般工程要求排水坡度不低于千分之三；而土地开发利用中则认为坡度大于25°就不宜开发。坡向则与光照有关，对建筑朝向和作物种植的选择非常重要。

坡度是表征地面上某点倾斜程度的一个量，是点位的函数。坡度有两种定义，其一为基于数学的定义：坡度是一个矢量，是空间曲面上某点的外向法线方向与垂直方向 Z 轴正向的夹角；其二为基于自然地理的定义：坡度是一个无量纲值，是空间曲面上某点

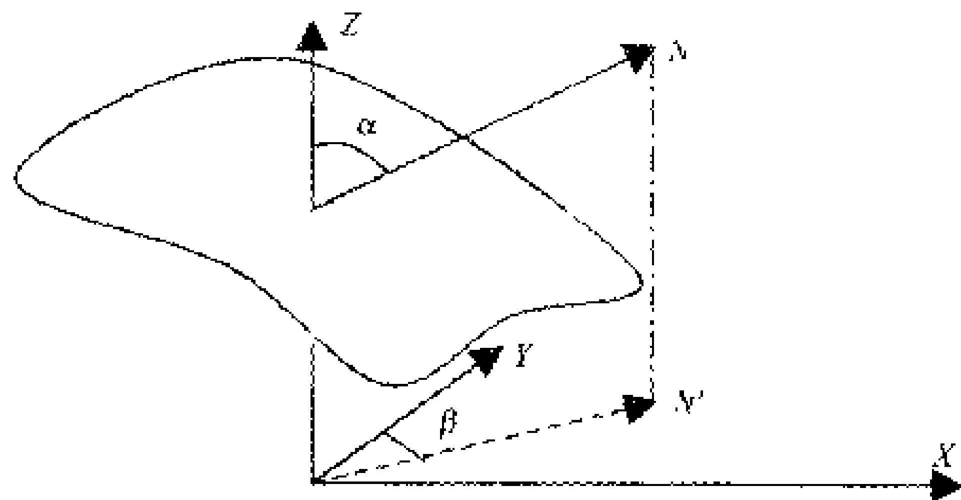


图 18-14　坡度坡向的定义

的切平面与水平面夹角的正切。通常，人们在使用过程中往往忽略坡度的矢量性，直接使用“坡度”这个无量纲的反正切来表示实际意义上坡度值 β（0～90°，水平面为 0°，立面为 90°）。坡向则定义为地表单元的外法线向量在水平面的投影线与地理北方向的夹角 α，即切平面上沿最大倾斜方向的矢量在水平面上的投影方向。如图 18-14 所示。

18.4.1 基于矢量的坡度坡向度量算法

原苏联著名的地图学家伏尔科夫于 20 世纪 50 年代提出了一种基于等高线计算地表坡度的算法：

$$\alpha = \arctan \frac{h \sum l}{P} \tag{18-24}$$

式中，h 为等高距；$\sum l$ 为测区内等高线总长；P 为测区面积。

显然，上式求出的是一个区域的平均坡度，且前提条件是测区内等高距相等。当测区较大或等高距不等时，用上式计算的误差较大。

为此，有人提出了一种基于统计学原理的变通办法。该法基于地形坡度越大等高线越密、地形坡度越小等高线越疏的特点，将测区划分为 $m \times n$ 个矩形子区，计算每一子区内等高线长度；然后根据回归分析技术统计出单位面积内等高线长度与坡度之间的回归模型；最后根据回归模型和已知的子区等高线长度，换算出子区的坡度值。该法的关键是建立单位面积内等高线长度与坡度之间的回归模型。其最大优点是可操作性强，且不受数据量的限制，适合处理海量数据。

如果某一地形表面可以用一个曲面函数 $Z = f(x, y)$ 进行拟合，则曲面上任一点 $P_0(x_0, y_0, z_0)$ 的切平面方程和外法线方程分别为：

$$\begin{aligned} Z &= Ax + By + C \\ &= f_x(x_0, y_0)x + f_y(x_0, y_0)y + C = 0 \end{aligned} \tag{18-25}$$

$$f_x^{-1}(x_0, y_0)(x - x_0) + f_y^{-1}(x_0, y_0)(y - y_0) + (z - z_0) = 0 \tag{18-26}$$

法线的方向数为 $A = f_x(x_0, y_0)$、$B = f_y(x_0, y_0)$ 和 $C = -1$（Z 轴的方向数为 $A = 0$，$B = 0$，$C = 1$）。故 $P_0(x_0, y_0, z_0)$ 点处坡度为：

$$\alpha = \arccos \frac{1}{\sqrt{f_x^2(x_0, y_0) + f_y^2(x_0, y_0) + 1}} \tag{18-27}$$

$P_0(x_0, y_0, z_0)$ 点处坡向为：

$$\beta = \arctan \frac{f_x(x_0, y_0)}{f_y(x_0, y_0)} \tag{18-28}$$

由于坡向在 $(0, 2\pi)$ 范围内变化，而按上式求得的坡向取值范围仅为 $(-\pi/2, \pi/2)$。故需要求该点外法线方向数的正负值组合，并来按表 18-1 取值。

表中“≈”表示 A（或 B）趋近于零的特殊情况。此时，坡向趋近于与 X 或 Y 坐标轴平行。

表 18-1 坡向的取值

A	>0	>0	>0	≈ 0	≈ 0	<0	<0	<0
B	>0	≈ 0	<0	>0	<0	>0	≈ 0	<0
α	$\left[0, \frac{\pi}{2}\right]$	$\frac{\pi}{2}$	$\left[-\frac{\pi}{2}, 0\right]$	0	0	$\left[-\frac{\pi}{2}, 0\right]$	$-\frac{\pi}{2}$	$\left[0, \frac{\pi}{2}\right]$
坡向 β	β	0	$2\pi - \beta$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	$\pi + \beta$	π	$\pi + \beta$

18.4.2 基于栅格的坡度坡向度量算法

基于规则格网 DEM 的坡度坡向度量将在第十九章中重点讨论。本节仅对基于 TIN 的坡度计算作一简单介绍。

TIN 上任意三角形可以用一平面方程 $Z = a_0 + a_1 x + a_2 y$ 表示，平面上坡度处处相等，可以用下式计算该三角形的坡度：

$$\alpha = \arccos \frac{1}{\sqrt{a_1^2 + a_2^2 + 1}} \tag{18-29}$$

对于一个区域而言，既可以将按上式计算的区域内各三角形的坡度求取总平均值(或按三角形面积进行加权平均)，也可以先用最小二乘逼近技术将区域拟合为一平面，然后将按上式计算的坡度作为其平均坡度。

参 考 文 献

龚健雅. 2001. 地理信息系统基础. 北京：科学出版社. 256～262

李志林，朱 庆. 2000. 数字高程模型. 武汉：武汉测绘科技大学出版社. 142～143

郭仁忠. 2000. 空间分析. 武汉：武汉测绘科技大学出版社. 153～160，193～194

第十九章　数字地形分析算法

地形分析是地形环境认知的一个重要手段。传统的地形分析是基于二维平面纸质地图进行的，需要大量的人工计算。DEM 技术和可视化技术的发展，使得地形分析的手段、效率、功能等发生了一次飞跃，三维可视或虚拟地理环境中的实时、交互模拟与分析也成为可能。地形数据的应用可以分为两类，即直接应用和间接应用。直接应用是面向基础测绘，将 DEM 本身作为测图自动化的重要组成部分和 GIS 数据库的基础；间接应用是面向专业用户，将 DEM 经过某种分析和变换产生各相关专业所需的派生产品。直接或间接应用过程中，从 DEM 所表达的地形属性中提取各类地形参数和特征因子的过程即为数字地形分析。

用来描述地形特征和空间分布的地形参数有很多，不同学科和领域对此的理解和分类也不尽相同。例如，按地学应用范畴可以分为一般地形属性和水文特征（Wood，1996）；按地形分析的复杂性，可以分为基本地形因子计算和复杂地形分析（李志林，朱庆，2000）；按地形要素的复杂性可以分为单要素参数和复合参数，其中单要素参数由高程数据直接得到，而复合参数是几个单要素参数的函数（Wilson J. P. and Gallant J. C.，2000；邬伦等，2001）；按地形参数的计算特性可以划分为局部地形参数和非局部地形参数（Florinsky I. V.，1998）；也有综合地形要素的关系特征和计算特征，分为单要素、复合要素和局部要素三类（刘学军，2002）。本章将综合相关知识，从基本地形分析、地形特征提取、典型应用分析三个方面来剖析有关算法。

§19.1　基本地形因子分析算法

本质上讲，DEM 是地形的一个数学模型，是一个或多个函数的集合。许多基本地形因子可从这些函数中进行一阶或二阶求导得出来的，也有的是通过某种组合或复合运算得到的。基本地形因子包括斜坡因子（坡度、坡向、坡度变化率、坡向变化率、日照强度）、面积因子（表面积、投影面积、剖面积）、体积因子（山体体积、挖填体积）和面元因子（面元相对高差、面元粗糙度、面元凹凸系数、高程变异等）。其中面积因子和体积因子作为空间度量的重要内容在第十八章已经介绍，不再赘述；基于矢量的坡度、坡向度量，也已在第十八章论及，本节不再涉及。因此，本节将主要基于规则格网 DEM 来剖析基本地形因子的其他有关分析算法。

19.1.1　斜坡因子分析算法

1. 坡度/坡向算法

基于 DEM 的坡度、坡向计算，人们进行了大量研究，提出了许多算法，其中最简单的是最大坡降法（Deterministic eight-node，D8）。最大坡降法的原理是利用中心格网

8 邻域的高程来计算坡度和坡向，其最大者即为该中心格网的坡度，所在方向即为该中心格网的坡向。该法虽然计算简单，计算效率也较高，但高程误差对计算结果影响较大 (Burrough P. A. and Mcdonnell R. A., 1998)。为了克服最大坡降法的缺点，人们研究并提出了多种改良方法。经证明，其中的拟合曲面法是求解坡度的最佳方法（李志林，朱庆，2000）。如图 19-1 所示，拟合曲面一般采用二次曲面，二次曲面的基本分割类似于规则 DEM 格网，即 3×3 规则窗口。每个窗口（0～8）中心为一个高程点（高程为 e_i），中心窗口的坡度 Sl、坡向 As 的计算公式如下：

$$Sl = \arctan\sqrt{Sl_x^2 + Sl_y^2} \tag{19-1}$$

$$As = 180° - \arctan(Sl_y/Sl_x) + 90° \times \frac{Sl_x}{|Sl_x|} \tag{19-2}$$

式中，Sl_x 为该点沿 X 方向的高程变化率；Sl_y 为该点沿 Y 方向的高程变化率。

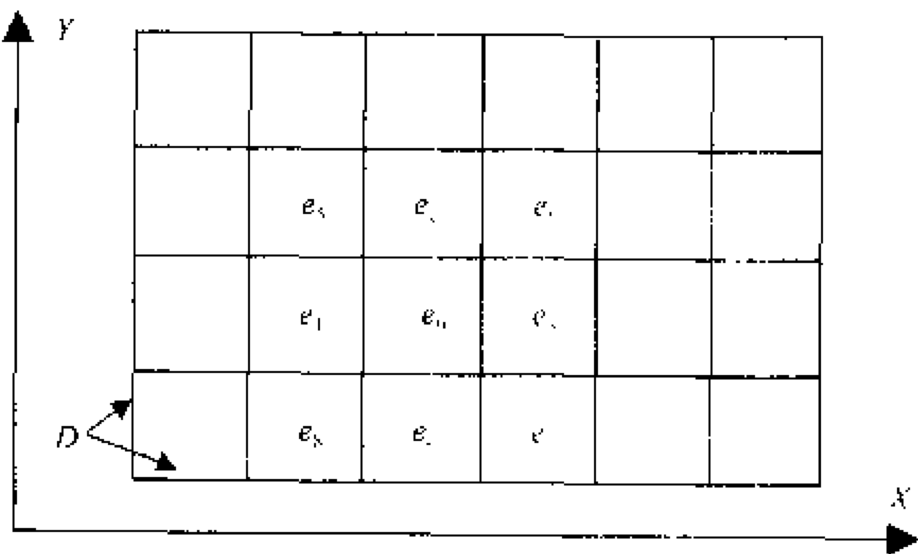


图 19-1　坡度/坡向的 3×3 窗口计算示例

由上可见，坡度坡向计算的关键是求解 Sl_x 和 Sl_y。刘学军将基于规则格网 DEM 中 3×3 移动窗口（包括基于 3×3 移动窗口的局部拟合曲面）的坡度坡向算法归纳为表 19-1。

基于格网 DEM 的各类算法中所采用的 f_x 和 f_y 的计算方法有以下 7 种：

算法 1（二阶差分法，矢量算法，不完全四次曲面法）

$$\begin{cases} Sl_x = \dfrac{e_1 - e_3}{2D} \\ Sl_y = \dfrac{e_4 - e_2}{2D} \end{cases} \tag{19-3}$$

算法 2（三阶不带权差分法，线形回归平面法，非带权限制及非限制二次曲面法）

$$\begin{cases} Sl_x = \dfrac{(e_8 + e_1 + e_5) - (e_7 + e_3 + e_6)}{8D} \\ Sl_y = \dfrac{(e_8 + e_4 + e_7) - (e_5 + e_2 + e_6)}{8D} \end{cases} \tag{19-4}$$

表 19-1 基于 3×3 移动窗口的坡度、坡向算法

（据李志林，朱庆，2000）

<table>
<tr><td rowspan="6">数值分析法</td><td colspan="3">简单差分法（Simple Difference）</td></tr>
<tr><td colspan="3">最大坡降法（Maximum Drop Slope）（O' Callaghan J. F., and Mark D. M., 1984）</td></tr>
<tr><td colspan="3">二阶差分法（2nd Finite Difference）（Fleming and Hoffer, 1979；Unwin, 1981）</td></tr>
<tr><td rowspan="2">三阶差分法（3rd Finite Difference）</td><td colspan="2">不带权（Non-weighted）（Sharpnack D. A., and Akin G., 1969）</td></tr>
<tr><td colspan="2">带权（Weighted）（Unwin, 1981）</td></tr>
<tr><td colspan="3">Frame 差分法（Frame Finite Difference）</td></tr>
<tr><td rowspan="7">局部曲面拟合法</td><td colspan="3">线形回归平面法（Sharpnack and Akin, 1969）</td></tr>
<tr><td rowspan="5">二次曲面法</td><td rowspan="2">限制型</td><td>带权（Wood J. D., 1996）</td></tr>
<tr><td>不带权（Wood, 1996）</td></tr>
<tr><td rowspan="2">非限制型</td><td>带权（Horn, 1981）</td></tr>
<tr><td>不带权（Evans, 1980）</td></tr>
<tr><td colspan="2">不完全四次曲面（Zevenbergen L. W., and Thorn C. R., 1987）</td></tr>
<tr><td colspan="4">空间矢量法（Ritter, 1987）</td></tr>
<tr><td colspan="4">快速傅立叶变换（Papo H. B. and Gelbman E., 1984）</td></tr>
</table>

算法 3（三阶反距离平方权差分法，带权限制及非限制二次曲面法）

$$\begin{cases} Sl_x = \dfrac{(e_8 + 2e_1 + e_5) - (e_7 + 2e_3 + e_6)}{8D} \\ Sl_y = \dfrac{(e_7 + 2e_4 + e_8) - (e_6 + 2e_2 + e_5)}{8D} \end{cases} \tag{19-5}$$

算法 4（三阶反距离权差分法）

$$\begin{cases} Sl_x = \dfrac{(e_8 + \sqrt{2}e_1 + e_5) - (e_7 + \sqrt{2}e_3 + e_6)}{8D} \\ Sl_y = \dfrac{(e_7 + \sqrt{2}e_4 + e_8) - (e_6 + \sqrt{2}e_2 + e_5)}{8D} \end{cases} \tag{19-6}$$

算法 5（Frame 差分法）

$$\begin{cases} Sl_x = \dfrac{(e_8 + e_5) - (e_7 + e_6)}{4D} \\ Sl_y = \dfrac{(e_8 + e_7) - (e_5 + e_6)}{4D} \end{cases} \tag{19-7}$$

算法 6（简单差分法）

$$\begin{cases} Sl_x = \dfrac{e_1 - e_0}{D} \\ Sl_y = \dfrac{e_4 - e_0}{D} \end{cases} \tag{19-8}$$

算法 7

$$\begin{cases} As = \arctan \dfrac{A}{B} \\ A = (e_5 + e_8) - (e_6 + e_7) \\ B = (e_7 + e_8) - (e_5 + e_6) \end{cases} \tag{19-9}$$

如图 19-2 所示的象限和方位分布，式（19-9）有如表 19-2 所示的各种可能结果。

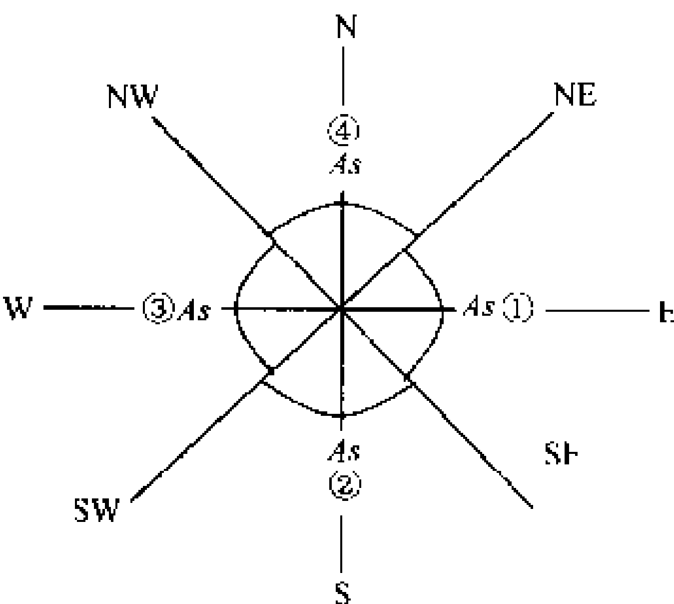


图 19-2　坡向与地理方位

表 19-2　坡向及其地理意义

A	B	As	地理坡向	俗称	所属区域
=0	=0	/	/	平地	/
<0	0	90°	E	半	①
<0	>0	|As|>45°	E-NW	阳	①
<0	<0	|As|>45°	E-SE	坡	①
=0	>0	/	S	阳	②
≠0	>0	|As|≤45°	SW-SE	坡	②
>0	=0	90°	W	半	③
>0	>0	|As|>45°	W-NW	阴	③
>0	<0	|As|>45°	W-SE	坡	③
=0	<0	/	N	阴	④
≠0	<0	|As|≤45°	NW-NE	坡	④

Hodgson M. E.(1995)对坡度计算的上述前 4 种方法进行了对比研究，得出结论：算法 1 的精度和效率均最高，其次是算法 3。而某些商用 GIS 软件所选择的坡度算法是不一样的，如 ERDAS Imagine 采用算法 2，ArcVIEW 采用算法 3。

2. 坡度变化率/坡向变化率算法

如图 19-1 所示，坡度变化率的定义为：任意格网点的坡度变化率等于该格网点的 8 个相邻格网点坡度变化率中绝对值最大的一个，并与它符号相同。对位于区域边角的格网点，其坡度变化率根据其相邻的 3 个格网点的坡度变化率来确定；位于边界边缘的格网点，其坡度变化率根据其相邻的 5 个格网点的坡度变化率来确定。

设 0～8 号格网点的坡度为 Sl_i（$i = 0 \sim 8$），记

$$S_i = \begin{cases} \dfrac{Sl_i - Sl_0}{D} & i = 1,2,3,4 \\ \dfrac{Sl_i - Sl_0}{\sqrt{2}D} & i = 5,6,7,8 \end{cases} \tag{19-10}$$

于是，3×3 窗口中心点的坡度变化率 S_0 为：

$$S_0 = \mathrm{SGN}_{S_{\max}} \mid S_{\max} \mid \tag{19-11}$$

式中：$\mathrm{SGN}_{S_{\max}}$ 为表示 S_0 的方向与 $S_{\max}$ 相同。

$$\mid S_{\max} \mid = \mathrm{MAX} \mid S_i \mid (i = 1 \sim 8)。$$

坡向变化率的定义与算法与此类似，只需将上述定义和公式中坡度换为坡向即可。

3. 日照强度算法

日照强度是综合反映某一位置坡度、坡向的综合参数。受太阳高度角和方位角的影响，日照强度与地理经纬度、季节和每日时间有关。日照强度 E 的计算公式如下：

$$E = T \cdot G \times [\sin H \cdot \cos Sl + \cos H \cdot \cos(B - As) \cdot \sin Sl] \tag{19-12}$$

式中：T 为大气透射率；G 为太阳常数，$G = 1.36 \times 10^6$ 尔格/（s·cm²）；H 为太阳高度角；B 为太阳方位角；Sl 为坡度；As 为坡向。

当 $E = 0$ 时，表示见不到阳光，为阴面。

19.1.2 面元因子分析算法

1. 面元相对高差

所谓面元，是指 DEM 的一个规则格网或一个三角形。面元相对高差即为面元各顶点中最高点与最低点之间的高差，记为 Δh，计算公式如下：

$$\Delta h = \mathrm{MAX}(h_i) - \mathrm{MIN}(h_i) \tag{19-13}$$

式中：h_i 为面元第 i 个顶点的高程（规则格网：$i = 1 \sim 4$；三角形：$i = 1 \sim 3$）。

注意一个类似概念，叫做面元相对高程 ΔH，它不同于面元相对高差，常易混淆。面元相对高程是指面元顶点的平均高程与所在区域或所在流域的最低高程 $H_{\min}$ 之差。

$$\Delta H = \frac{1}{4}\sum_{i=1}^{4} e_i - H_{\min} \tag{19-14}$$

2. 面元粗糙度

面元粗糙度是指面元表面积 S 与其水平投影面积 S_p 之比，记为 C_Z，计算公式如下：

$$C_Z = \frac{S}{S_p} \tag{19-15}$$

当 $C_Z = 1$ 时，粗糙度最小，表示该面元为水平面。

此算法有一个致命的问题，即对于实际光滑的斜平面，由于其斜度不同，将产生不同的粗糙度。显然，这一问题的产生，是由于粗糙度这一相对式定义不合理。

可以采用另外一种绝对式命题，即以面元的两条对角线的中点的高程差 D 的绝对值来表示面元的粗糙度：

$$D = \left| \frac{e_5 + e_7}{2} - \frac{e_6 + e_8}{2} \right| \tag{19-16}$$

D 越大，表示面元越粗糙；反之，表示面元越平坦。

3. 面元凹凸系数

称面元的最大高程顶点与其对角点的连线为主轴，主轴两端点高程平均值与该面元顶点的平均高程的比值，即为面元的凹凸系数，记为 C_D，计算公式如下：

$$C_D = \frac{h_{\max} + h'_{\max}}{2h_{\text{ave}}} \tag{19-17}$$

式中：h_{max}为面元顶点的最大高程；h'_{max}为面元最大高程顶点的对角点的高程；h_{eve}为面元顶点的平均高程。

当 $C_D>1$ 时，表示该面元为凸面；当 $C_D=1$ 时，表示该面元为平面；当 $C_D<1$，表示该面元为凹面。

4．面元高程变异系数

面元高程变异系数是指面元的4顶点的高程标准差 σ 与其平均高程 $\bar{h}$ 的比值。

$$\begin{cases} s = \dfrac{\sigma}{\bar{h}} \\ \bar{h} = \dfrac{1}{4}\sum_{i=1}^{4} h_i \\ \sigma = \sqrt{\dfrac{1}{4}\sum_{i=1}^{4}(h_i - \bar{h})^2} \end{cases} \tag{19-18}$$

§19.2 地形特征提取算法

地形特征是指描述地形形态的地表点、线和面所构成的地形起伏变化的骨架。地形特征点包括脊点、谷点和鞍点等；脊点、谷点分别是指区域内相对高程最高和最低的点集；地形特征线包括山脊线、山谷线，鞍点即是山脊线与山谷线交点；地形面状特征包括地面的凹凸性。近年来，随着GIS的应用和技术的发展，关于地形特征自动提取问题的研究十分活跃，对于扩展GIS的应用功能具有特别意义。从规则格网、TIN和等高线提取地形特征，是当前研究与应用的主要方面。

19.2.1 基于规则格网DEM的算法

规则格网DEM是一种十分重要的数据结构，许多国家的DEM数据都是以规则格网的数据矩阵形式提供的（如美国USGS提供分辨率为30m、90m的DEM；澳大利亚、日本、英国均提供50m分辨率的DEM；而加拿大为93m的DEM，法国为100m的DEM）。可见，目前大多数的地形特征提取算法都是基于规则格网的。算法的大致原理可以归结为：

- 直接计算法；
- 基于图像处理技术；
- 基于几何分析原理；
- 基于流水物理模拟分析原理；
- 基于几何分析与物理模拟分析相结合的原理。

1．直接计算法

如图19-1所示，通过对某格网及其四邻格网的高程比较，就可以直接判断该格网是脊点、谷点还是鞍点。结果如表19-3所示。

表 19-3　脊点、谷点还是鞍点的直接算法

$(e_1-e_0)\times(e_3-e_0)$	(e_1-e_0)	地形特征点性质
>0	>0	谷点
	<0	脊点
$(e_2-e_0)\times(e_4-e_0)$	(e_2-e_0)	鞍点
>0	>0	谷点
	<0	脊点

在脊点、谷点和鞍点提取的基础上，还可以进一步计算区域的沟谷长度、沟谷密度和沟谷深度。沟谷长度是指区域内谷点的总延伸长度；沟谷密度是指区域内沟谷总长度与区域面积的比值；而沟谷深度则是指某一区域内若干个谷点切割深度的平均值。此处，谷点切割深度是指该谷点与离其最近的脊点之间的高差。

2．基于图像处理技术

由于规则格网 DEM 可以看成为栅格，因而可以借用基于栅格的图像处理技术来提取地形特征。提取过程分两步：首先提取地形特征点，然后将特征点连成地形特征线。

Peucker T. K. 和 Douglas D. H. 于 1975 年提出了一种简单的特征点提取算法，称为移动窗口法。该法的思路是用一个 2×2 的窗口对 DEM 格网依次扫描，将窗口中的最低高程点作标记，扫描结束后，凡是自始至终未作标记的点即为峰点；反之，用一个 2×2 的窗口对 DEM 格网依次扫描，将窗口中的最高高程点作标记，扫描结束后，凡是自始至终未作标记点即为谷点。

Band L. E. 于 1986 年对此算法进行了改进，采用表示河流段上下游顶点并进行细化处理的方法将所有细碎断裂的线连接起来，形成地形特征线。此外，还有一些其他的提取地形特征线的方法，如 Andrew 与 Skidmore 算法（Skidmore A. K.，1989）等。限于篇幅，不再赘述。

3．基于几何分析原理

基于几何分析的一个典型算法是断面极值法。该法的基本思想是认为地形断面曲线上的极大值点为分水点，而极小值点为汇水点。其实施过程是：先找出 DEM 的系列纵横断面上的极大、极小值点，作为地形特征线上的备选点；然后根据一定的条件或准则，将候选特征点划归所属的地形特征线。

应用表明，该法有两点缺陷：

（1）由于在确定地形特征线时，全区域采用同一个曲率阈值作为判定地形特征线上点的条件，故忽略了每条特征线必然存在的曲率变化现象。当阈值选择较大时，会丢失许多特征线上的点，导致后续跟踪的特征线较短且存在间断；当阈值选择较小时，会将许多本来不是地形特征线上的点误认为是，这将给后续地形特征线的跟踪带来麻烦。

（2）由于该法仅选择纵、横两个断面来确定其高程变化的极值点，因此，所确定的

地形特征线带有一定的近似性，有时会遗漏某些地形特征线。为改良这一缺陷，有学者（如黄培之，1990）提出在规则格网的数字地面模型数据中增加规则格网对角线方向上的一组断面，有助于克服上述缺陷。

4. 基于流水物理模拟分析原理

基于流水物理模拟分析的基本思想是：按照流水从高至低的自然规律，顺序计算每一地形点的汇水量，然后按汇水量单调增加的顺序，由高到低找出区域中的每一条汇水线。根据已经得到的汇水线，通过计算就可以找出各自汇水区域的边界线，此边界线即为分水线。

应用表明，该法也有两点缺陷：

（1）由于该算法所计算的汇水量与高程有关，计算所得必然是高程大的地形特征线上的点汇水量小，而高程小的地形特征线上的点汇水量大。因此，可能导致位于低位的非地形特征点的汇水量也较大而被当成地形特征线上的点；而位于高处的地形特征线上的点则因为汇水量小而被排除；这就造成用该算法所确定的地形特征线在两端的效果很差。

（2）由于该算法将各汇水区域的公共边界视为分水线，因此，它所确定的分水线均为闭合曲线，这也与地形实际变化不相符。

5. 基于几何分析与物理模拟相结合的原理

鉴于基于几何分析和基于流水物理模拟的算法均存在缺陷，有学者提出将两者结合起来，取长补短。其基本思路是：首先采用较稀疏的DEM格网数据，按流水物理模拟算法提取区域内概略的地形特征线；然后用其作引导，在其周围邻近区域对地形进行几何分析，进而精确地确定区域内的地形特征线。

该算法的关键为：求解已提取的概略地形特征线与DEM格网线的交点，在交点附近的小区域内，对DEM数据进行几何分析，找出该区域内与概略地形特征线正交方向的地形特征线的精确位置。该法操作的整个过程分5步：① 概略DEM建立；② 地形流水物理模拟；③ 概略地形特征线提取；④ 地形几何分析；⑤ 地形特征线的精确确定。

19.2.2 基于等高线和TIN的算法

TIN和等高线是DEM的另外两种数据结构形式。从TIN和等高线中提取地形特征的算法有多种，包括等高线曲率判别法、等高线垂线跟踪法、等高线骨架法和Voronoi图骨架法等。

1. 等高线曲率判别法

等高线曲率判别法（费立凡，1994）的主要步骤为：① 首先计算每条等高线上一定间距的离散点的曲率绝对值；② 将曲率绝对值大于某一给定阈值的离散点选择出来，作为地形特征线的备选点；③ 每条等高线均按上述两步处理完之后，找出区域内的峰点和谷点，并以这些点为起始点，根据一定的搜索策略将已确定的备选点纳入各自所在的山脊线和山谷线。

显然，该法与基于几何分析原理的算法一样，均将地形特征点的判定和该点地形所属地形特征线的确定分开考虑，因此，存在与几何分析算法同样的阈值缺陷问题。此外，在地形复杂（地形破碎）和等高线不光滑（存在噪音）的区域，地形特征线的跟踪十分困难。

2. 等高线垂线跟踪法

等高线垂线跟踪法的过程是：① 以一定的步长在数字化等高线上选取离散点，并计算各离散点处等高线的法线方向的单位矢量；② 将该单位矢量在 X、Y 方向进行分解；③ 以此为基础，通过内插计算得到其他非等高线上的点（所需点）所在处的等高线的法线方向矢量；④ 按高程由高到低的顺序进行等高线法向矢量跟踪，并求取跟踪所得的各等高线法线轨迹的交汇点；⑤ 交汇点的有序连接即为地形特征线。

该算法的实质是找出地形曲面上每点的流水轨迹，这些轨迹的顺序连接即为汇水线，各汇水线之起点所组成的线即为流水线。该法与等高线曲率法相比有抗干扰（噪音）能力强的优点，但计算工作量大。

3. 等高线骨架法

骨架法又称中心轴化法，近年来被广泛用于图像、图形处理领域。在图形图像处理领域，图形骨架是指二维图形边界内距其两侧边界等距离的点所组合成的线。在 GIS 领域，等高线骨架法的实质是将同一条等高线的中心轴线视为地形特征线，即认为地形特征线两侧的地形为对称变化。显然，这与大多数地形变化状况是不相符的。因此，用该法所得的地形特征线带有很大的近似性。

4. Voronoi 图骨架法

如果基于等高线先生成 Delauny TIN，则 Delauny TIN 所对应的 Voronoi 多边形的各个顶点即构成地形的骨架点。这些骨架点可能构成三种不同的特征形态（如图 19-3 所示）：一为骨架线或中心轴线，二为地形特征线，三为小毛刺。其中中心轴线上点的

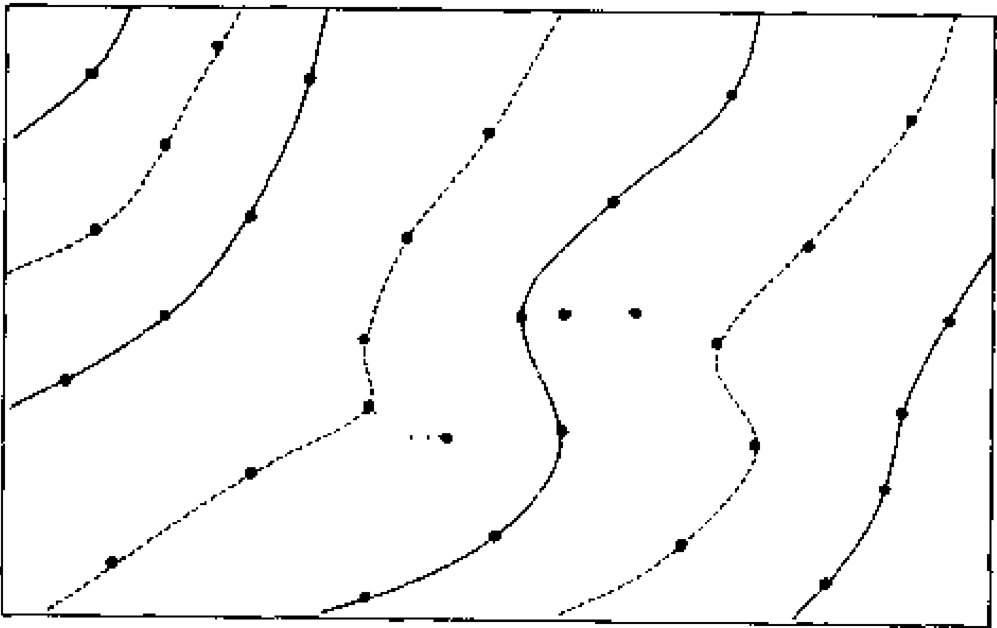


图 19-3 由 Voronoi 多边形顶点构成的骨架
（据李志林，朱庆，2000）
黑线为等高线；虚线为中心轴线；点线为地形特征线

高程是两侧等高线高程的平均值；而地形特征线上点的高程则可以通过内插得到；毛刺产生的原因主要是等高线不光滑、有许多弯曲等。

Voronoi 图骨架法的最大优点是，地形骨架线和地形特征线在建立 Delaunay TIN 和 Voronoi 图的过程中可以直接得到。

§19.3 数字地形典型应用算法

19.3.1 流水分析算法

实际上，地形的水文特征是与地形特征密切相关的。例如，山脊线具有分水性，山谷线具有汇水性。所不同的是，水文分析还要在地形特征分析的基础上分析水系的流域范围、洪水淹没范围、污染影响范围、土壤侵蚀过程等。上述过程及现象研究的基础是流水模拟或溢水分析，也有人称之为路径分析算法（routing algorithm）（Desmet P. J., and Govers G., 1996a；刘学军，2002）。可以基于等高线、TIN 和规则格网模型来进行流水分析。基于等高线和 TIN 的分析相对简单些，因为流水线实际上是与等高线垂直的，而流径方向取决于 TIN 的地形位置。但由于规则格网是 DEM 的主要形式，我们主要讨论基于规则格网 DEM 的流水分析算法。

由于地表本身及流水过程的双重复杂性，流水算法均不同程度地以各种假设（如 D8 算法假设水的流向是 8 个可选方向之一）为基础。迄今为止，基于格网 DEM 的算法如表 19-4 所示。由于算法的前提假设不同，往往导致分析结果不同，甚至差异很大。

表 19-4 基于格网 DEM 的流水算法

单流向算法 SFD	最大坡降算法 D8（Deterministic eight node）（Mark D. M., 1984；O'Callaghan J. F. and Mark D. M., 1984）
	随机四方向算法 Rho4（Random four-node）（Fairefield J. and Leymarie P., 1991）
	随机八方向算法 Rho8（Random eight-node）（Fairefield J. and Leymarie P., 1991）
	流向驱动法（Aspect-Driven）（Lea E. J. et al, 1992；Gardner T. W. et al., 1990）
多流向算法 MFD	基于坡度的 MFD（Multiple Directions based on Slope Algorithm）（Quinn P. F. et al., 1991）
	基于坡度指数的 MFD（Freeman T. G., 1991）
	流管法 DEMON（Costa-Cabral M. C. and Burges S. J., 1994）
	无穷方向算法 Dinf（Tarboton D. G., 1997）
	形态算法（Pileojo et al., 1998）
多级骨架算法（Meisels A. et al., 1995）	

单流向算法与多流向算法的原理差异如图 19-4 所示。

此外，洼地是影响流水过程的重要因素。自然条件下，水往低处流，一旦遇到洼地，那么水流总是先将洼地填满，然后再从该洼地的某一最低出口流出，继续其流动过程。但如果不考虑洼地集水过程对溢水过程的动态影响（比如推迟溢洪蔓延），可以先按照某种规则对区域 DEM 进行无洼地化处理（李志林、朱庆，2000），以此来简化和加快流水方向和流水量的分析计算。

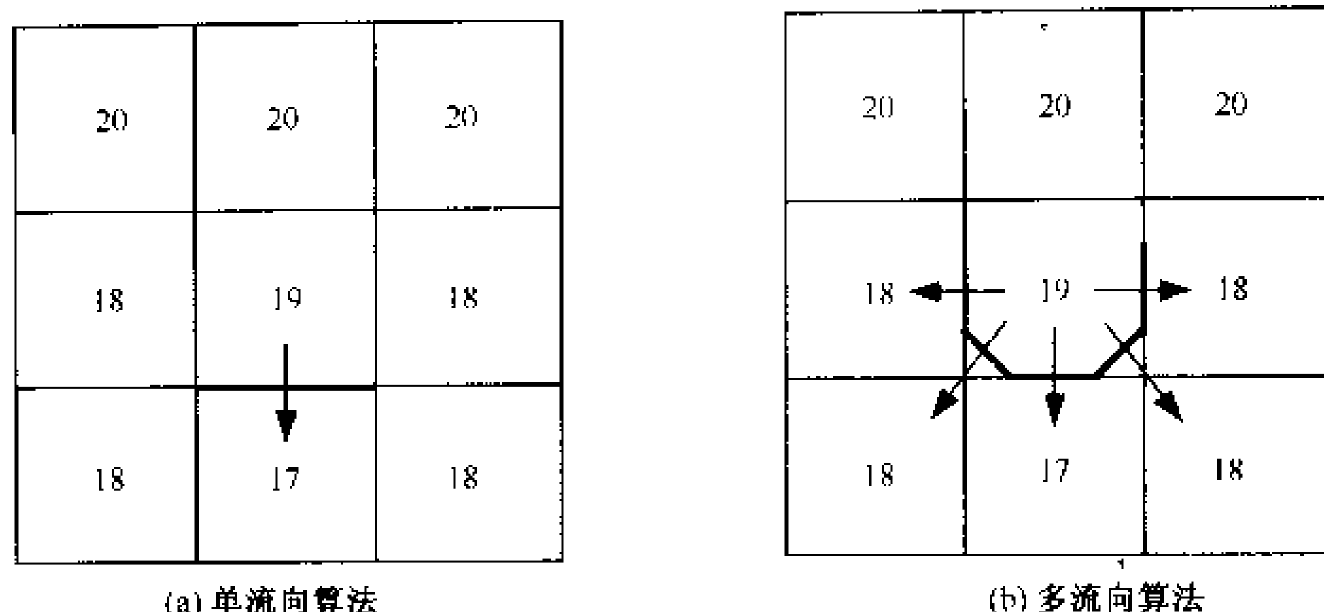


图 19-4 单流向算法与多流向算法的原理比较

（据 Wolock D. M. and McCabe G. J., 1995）

黑线为流向宽度；箭头为流向；数字为格网高程值

在格网 DEM 上进行流水模拟分析包括两项基本内容，即格网单元的水流方向确定和格网单元向较低单元溢水的流量分配。

1. 水流方向矩阵计算

对于 DEM 格网，水流方向是指水流离开此格网时的指向（李志林，朱庆，2000）。单流向算法 D8 的致命弱点是其流向的确定性，即一律选择最陡坡降方向为中心格网的流向；而 Rho4 和 Rho8 是 D8 的改进版本，其共同点在于流量比例的分配：即下游格网接受上游格网的全部流量，但通过引入一个随机概率参数从所有较低单元中选择一点为中心格网流向。

多流向算法则认为水流分布具有分散性质，即水流方向具有不确定性，水流可能同时流向该格网的 8 个邻域中某几个邻域格网。不同的多流向算法主要体现在流量分配的比例不同。

Quinn P. F. et al.（1991）按坡度和流向宽度关系分配中心格网流入下游格网的流量公式为：

$$F_i = \frac{L_i \tan\beta_i}{\sum_{i=1}^{n} L_i \tan\beta_i} \qquad (\tan\beta_i > 0, n \leqslant 8) \qquad (19\text{-}19)$$

式中，β_i 为中心格网到下游格网 i 的坡度；L_i 为流向宽度（当位于对角线方向时，L_i 为$\sqrt{2}/4$倍格网间距；反之，L_i 为 1/2 倍格网间距）。

Freeman T. G.（1991）的流量分配公式为：

$$F_i = \frac{(\tan\beta_i)^p}{\sum_{i=1}^{n} (\tan\beta_i)^p} \qquad (\tan\beta_i > 0, n \leqslant 8) \qquad (19\text{-}20)$$

式中，p 为坡度指数，通过圆锥实验表明，$p=1.1$ 比较合适。

为了统计方便，以一个确定的数字代表某栅格的水流方向的不确定性，使得后续处理过程中可以从这一确定的数字推知栅格的水流方向和其邻域栅格状态，水流方向值可

以用 2 的幂值来表示，如图 19-5 所示。

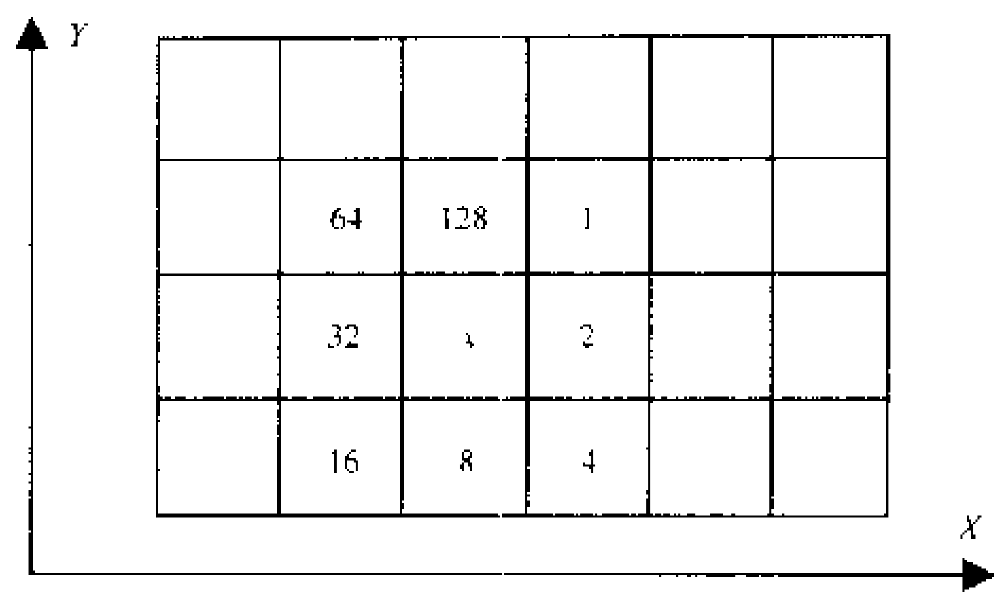


图 19-5　水流方向的数字化定义

此外，在分析水流方向时，还要考虑中心栅格与邻域栅格的距离，即格网之间的距离与邻域方向有关。当邻域对中心栅格的方向值为 2 (2^1)、8 (2^3)、32 (2^5)、128 (2^7) 时，距离为一个单位；为 1 (2^0)、4 (2^2)、16 (2^4)、64 (2^6) 时，距离为$\sqrt{2}$个单位。据此，可以定义距离权落差概念为：中心栅格与邻域栅格之间的落差（中心栅格的高程减去邻域栅格的高程）除以格网间距。水流方向确定的一般步骤为：

(1) 对区内 DEM 边缘格网，均赋予指向其外部边缘的方向值；

(2) 对区内 DEM 内部格网（非边缘格网），计算其 8 邻域的距离权落差 Δh_i ($i=1\sim8$)；

(3) 确定具有最大距离权落差的格网：① 如果 Δh_i ($i=1\sim8$) 均<0，则赋予负值，表示该格网方向未定（经洼地填充后不会出现）；② 如果 Δh_i ($i=1\sim8$) 均$\geqslant0$，且最大值只有一个，则对应此最大值得邻域栅格的方向值即为本中心格网的流水方向值；③ 如果 Δh_i ($i=1\sim8$) 均>0，且最大值有不止一个，则以查表方式确定：若一侧的三邻域格网点有相同的距离权落差，则该侧的中间栅格方向为水流方向；若两侧有两个邻域格网点距离权落差相同，则任选一个格网，以其方向作为本中心格网的流水方向值。

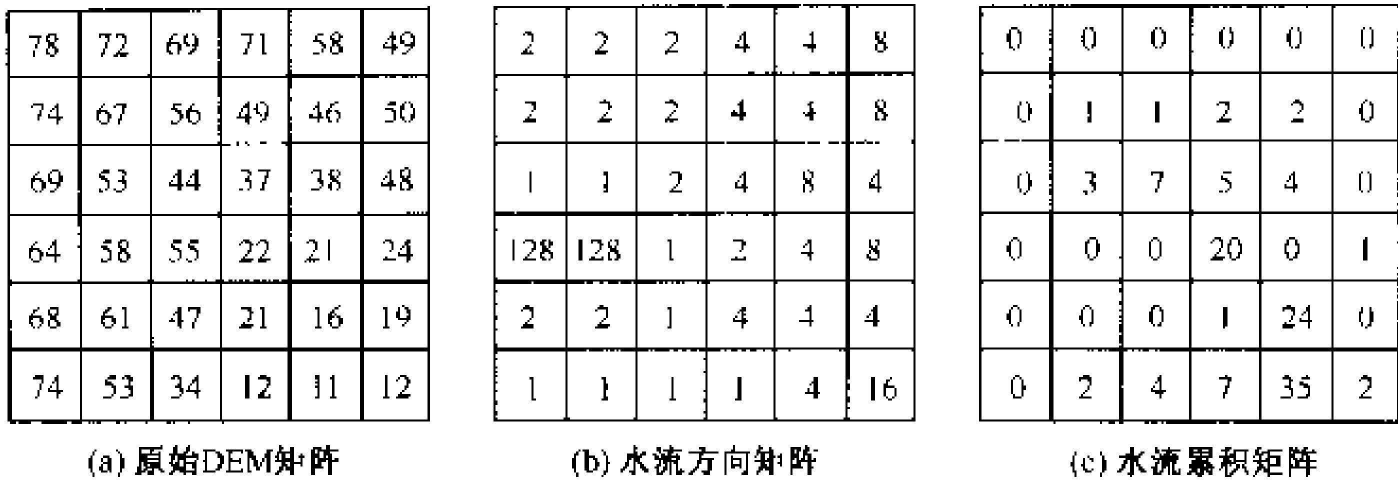
2. 水流量累积矩阵的计算

区域流水量累积矩阵值表示区域地形每点的流水累积量，它可以用区域地形曲面的水流模拟方法获得。水流模拟可以基于水流方向矩阵得到。其基本思想是：认为以规则格网表示的 DEM 的每一格网处有一个单位的水量，根据水往低处流的自然规律，基于水流方向矩阵计算每一格网处流过的水量。在此过程中，使用了权值全为 1 的权矩阵；应用中，如果考虑到降水不均匀等特殊情况，可以使用特定的权矩阵，以便更精确地计算水流累积值。图 19-6 所示为一个计算实例的结果。

上述算法是传统水文分析模型的基础算法，已在一些商业软件中实现。实际使用中，该算法仍然存在一些问题有待改进：

(1) 计算复杂，计算量大，且计算量随 DEM 格网规模增大或尺寸减小而成倍增加；

(2) 由于假设了流水的 8 邻域方向的强制选择性，与自然流水的多样性和随机性不相符合。有人提出若以 8 个方向的梯度为基础按比例分配从格网流出的水流的话，在非平坦地区的模拟结果可能更好。



78	72	69	71	58	49
74	67	56	49	46	50
69	53	44	37	38	48
64	58	55	22	21	24
68	61	47	21	16	19
74	53	34	12	11	12

(a) 原始DEM矩阵

2	2	2	4	4	8
2	2	2	4	4	8
1	1	2	4	8	4
128	128	1	2	4	8
2	2	1	4	4	4
1	1	1	1	4	16

(b) 水流方向矩阵

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	2

(c) 水流累积矩阵

图 19-6 洼水分析结果示例

（据李志林，朱庆，2000）

在流水方向和流水累积量分析的基础上，可以进一步进行单位汇水面积（specific contribution area，SCA）、总汇水面积（total contribution area，TCA）、流域与子流域范围（drainage and sub-drainage range）、流域网络（drainage network）、陆地流水路径（land water routine）等专业性水文问题的模拟与分析。刘学军（2002）的研究表明，D8 和 Rho8 不宜用来进行与 SCA 和 TCA 有关的水文参数计算；DEMON 算法以二维流管方式在格网结构上实现了基于等高线的 SCA 计算原理，比较符合水流的自然特性；Dinf 算法通过坡向引入流量比例分配参数，其计算精度比 DEMON 算法差但优于 FMFD 算法；而 FMFD 算法完全通过坡度确定流量分配比例，虽然符合水流的自然发散特性，但流量比例分配参数成为控制 SCA 精度的主要因素。

19.3.2 通视分析算法

通视分析也称可视性分析和可见性两点之间的通视性（intervisibility）分析及某点的可视域（viewshed）分析两类，属于基于地形的工程设计与布局优化范畴。应用领域包括雷达站、电台、通信站、导航站、哨所、瞭望塔、炮兵阵地、电子对抗基地等的设置，道路、电力线路、通信线路的选择等，以及各类可视查询与可视结构计算。

Nagy（1994）将基于 DEM 的可视性应用分为 3 个方面：

(1) 观察点问题：即在地形环境中选择数量最少的观察点，使得地形环境中的任意一个位置均至少有一个观察点与之通视。具体应用包括雷达站、电台、通信站、导航站、哨所、瞭望塔的定位布设。作为该问题的延伸，还有给定某一观察点，确定其最大可视范围（即可视域分析）；以及给定高程，查找能通视整个区域的观察点；或给定区域，查找能通视整个区域的最低观测点，具体应用如森林烽火塔、电视塔、旅游塔的定位等。

(2) 视线通信问题：即对于一个给定的两个或多个点，找到一个可视网络（visibility network），使得可视网络中任意两个相邻点之间可视。具体应用包括微波站、广播

电台、数字数据传输站点等网络的设计。

（3）表面路径问题：即指地形环境中与通视相关的路径设计问题。如预设两个点及一个观测点，从该两点之间的所有可能路径中找出一条观测点始终不可见的隐蔽路径；或预设若干点和一个起点，找出从起点出发遍历所有点的最短路径。其具体应用包括隐蔽通道、侦察路线、旅游路线、电力网络的设计等。

1. 通视性分析

两点通视性分析的算法有“剖面法”和“射线追踪法”两种。前者不仅适合基于规则格网、TIN 的 DEM，还适合基于等高线的 DEM；而后者仅适合基于规则格网和 TIN 的 DEM 。

“剖面法”的基本步骤为：

（1）确定通过两点并与 XY 平面垂直的剖面 S；

（2）求出地形模型中与剖面 S 相交的所有 DEM 格网边；

（3）判断相交的格网边是否位于两点连线之上，如果有至少一条边在其上，则两点不可通视；否则，通视.

“射线追踪法”的基本思路是：从其中一点开始，沿两点连线（射线）方向计算地形模型中与连线相交的第一个面元。如果该面元存在，则两点不可视；否则，可视。该法还可用于观测点的水平可视范围（射线为水平方向）计算和空间搜索范围（射线为 $-90°\sim+90°$）计算。

实际应用中，往往要求考虑各类地面建（构）筑物的高度、树冠高度及观测点仪器的高度（或人的身高）等因素对通视性的影响。此时，通视算法需作修正。如图 19-7 所示，以树冠遮挡影响为例，树冠遮挡距离计算公式为：

$$D_2 = \frac{A_3 - (A_1 + h_1)}{(A_2 + h_2) - (A_1 + h_1)} \times D_1 \qquad (19\text{-}21)$$

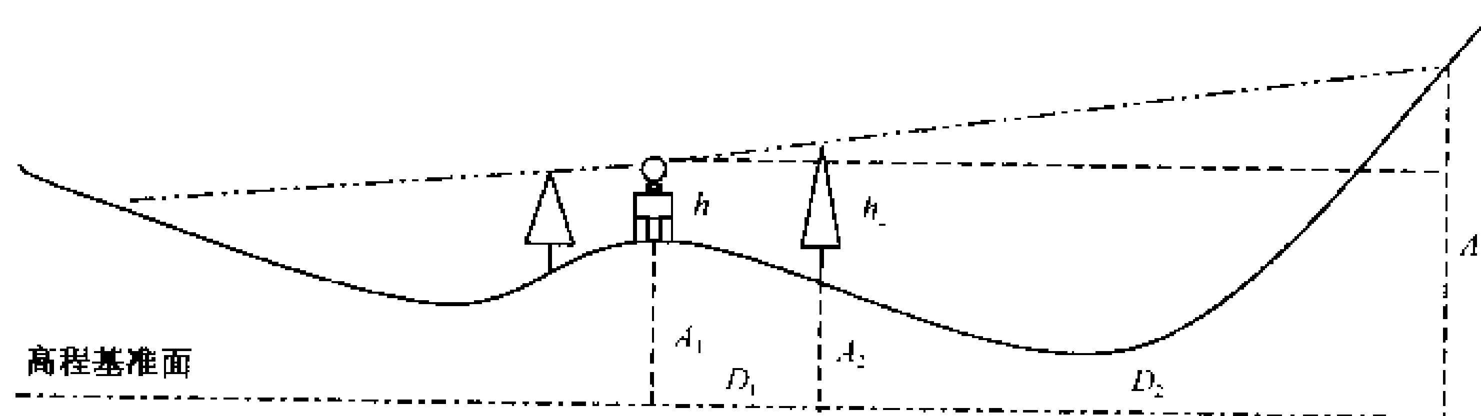


图 19-7 通视计算示例

式中：D_1 为观测点至树障的距离；D_2 为可视点至树障的距离，即树冠遮挡距离；A_1 为观测点高程；A_2 为树冠高程；A_3 为可视点高程；h_1 为观测者身高（或仪器高）；h_2 为树冠高度。

2. 可视域分析

基于规则格网可视域算法和基于 TIN 的可视域算法不太一样。在规则格网中，可

视域通常是以每个格网点的可视与不可视的离散形式表示的，称为“可视矩阵”。

基于规则格网的可视域算法的基本思路是：DEM 中的任一格网，将其与视点相连，判断其连线是否与 DEM 其他格网相交。若不相交，则该格网可视；否则，不可视。显然，这一算法存在大量的冗余计算。改进的方法有多种，比如首先判断连线与 DEM 的其他格网是否相交，若是，则离视点最近的交点网格为可视，沿连线方向该格网之后的其他格网均不可视，这样就可以减少不必要的计算次数。

基于 TIN 的 DEM 的可视域分析算法与三维场景中的隐藏面消除算法相似。在最复杂的新情况下，其时间复杂度为 $O(n^2)$。围绕该法，人们提出了多种改良措施来提高速度，读者可查阅计算几何和图形学中的有关文献。

参 考 文 献

黄培之．1990．关于坡度几个问题的研究．测绘学报，19（3）：225～228

李志林，朱庆．2000．数字高程模型．武汉：武汉测绘科技大学出版社．140～162

刘学军．2002．基于规则格网数字高程模型解译算法误差分析与评价．武汉大学博士学位论文．53～140

黄杏元，汤勤．1989．地理信息系统概论．北京：高等教育出版社．107～121

Band L. E．，1986．Topographic partition of watershed with digital elevation models．Water Resources Research，22（1）：15～24

Burrough P．A．，and Mcdonnell R．A．，1998．Principles of Geographical Information System．Oxford，UK：Oxford University Press

Costa-Cabral M．C．and Burges S．J．，1994．Digital elevation model networks（DEMON）：A model of flow over hillslopes for computation of contributing and dispersal areas．Water Resources Research，30（6）：1681～1692

Desmet P．J．J．and Govers G．，1996．Comparison of routing algorithms for digital Elevation models and their implication for predicting ephemeral gullies．IJGIS，10（10）：311～331

Freeman T．G．，1991．Calculating catchment area with divergent flow based on a Regular grid．Computer & Geosciences，17（3）：413～422

Fairfield J．and Leymarie P．，1991．Drainage networks from grid elevation models．Water Resources Research，27（5）：709～717

Florinsky I．V．，1998a．Combined analysis of digital terrain models and remotely sensed data in landscape investigations．Progress in Physical Geography，22（1）：33～60

Gandner T．W．，Sasowsky K．C．and Day R．L．，1990．Automated extraction of geomorphometric properties from digital elevation data．Zeitschrift for Geographical Information Science，12（8）：875～890

Hodgson M．E．，1995．What cell size does the computed slope/aspect angle represent？Photogrammetric Engineering and Remote Sensing，61：513～517

Lee J．，Snyder P．K．and Fisher P．F．，1992．Modeling the effect of data errors on feature extraction from digital elevation models．Photogrammetric Engineering and Remote Sensing，58（10）：1461～1467

Mark D．M．，1984．Automatic detection of drainage networks from digital elevation models．Cartographica，21（2/3）：168～178

Meisels A．，Raizman S．and Karmeli A．，1995．Skeletonizing a DEM into a drainage network．Computer & Geosciences，21（1）：187～196

O'Callaghan J．F．and Mark D．M．，1984．The extraction of drainage networks from digital elevation data．Computer Vision，Graphics，and Image Processing，28：323～344

Papo．H．B．and Gelbman E．，1984．Digital terrain models for slopes and curvatures．Photogrammetric Engineering and Remote Sensing，50（6）：695～701

Peucker T．K．and Douglas D．H．，1975．Detection of surface specific points by local parallel of discrete terrain elevation

data. Computer and Image Processing, (4): 375 - 387

Pilesjo P., Zhou Q. and Harrie L., 1998. Estimating flow distribution over digital elevation models using a form-based algorithm. Geographical Information Science, 4 (1-2): 44 - 51

Quinn P. F., Beven K., Chevallier P. and Planchon O., 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. In Beven K. J. and Moore I. D. eds, Terrain Analysis And Distributed Modelling in DEMs, Hydrology, John Willy & Sons, Chichester, UK. 63 - 83

Sharpnack D. A. and Akin G., 1969. An algorithm for computing slope and aspect from elevations. Photogrammetric Survey, 35: 247 - 248

Skidmore A. K., 1989. A comparison of techniques for the calculation of gradient and aspect from a grided digital elevation model. IJGIS, 3: 323 - 334

Tarboton D. G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research, 32 (2): 309 - 319

Wilson J. P. and Gallant J. C., 2000. Terrain analysis: principles and applications. John Wiley & Sons, UK

Wolock D. M. and McCabe G. J., 1995. Comparison of single and multiple flow Direction algorithms for computing topographic parameters in TOPMODEL. Water Resources Research, 31 (5): 1315 - 1324

Wood J. D., 1996. The geomorphological characterization of digital elevation model. Ph. D Thesis, University of Leicester

Zevenbergen L. W. and Thorne C. R., 1987. Quantitative analysis of land surface topography. Earth Surface Processes And Landforms, 12: 47 - 56

第二十章　空间统计分析算法

空间统计分析是GIS中的一项重要的特色工作，主要基于空间数据进行空间和非空间数据的分类、统计、分析和综合评价。空间统计分析方法很多，除一般的统计图表分析、密度分析之外，还有多变量统计分析（含主成分分析、主因子分析、关键变量分析、变量聚类分析和采样点聚类分析等）、空间分类分析（空间聚类分析、空间聚合分析和判别分析）以及层次分析等。

§20.1　多变量统计分析算法

随着数据采集技术的进步和采集手段的多样化，在同一采样点（或称样本点，数据点）上往往可以收集到几十种不同数据或变量，不仅给GIS模型的构建带来很大困难，也增加了数据库存储和系统运算的负担。从空间统计学和地理学的角度，这些数据或变量之间往往是相互关联的，只是关联的程度不同而已。如何从众多的变量中，找出一组相互独立的变量，使原始采样数据得以简化，是一个变量筛分的过程，此即多变量统计分析的主要任务。常用的多变量统计分析算法主要有主成分分析、主因子分析、关键变量分析和变量聚类分析（含采样点聚类分析）。

20.1.1　主成分与主因子分析算法

主成分分析是基于数理统计分析，求得各变量之间线性关系的表达式，进而将众多变量的信息压缩表达成若干具有代表性的合成变量，为空间聚类分析和应用模型构建铺平道路。设有 n 个采样点，每个采样点有 m 个变量，采样数据集合 X 的矩阵表示为：

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{20-1}$$

若将原始数据转换为一组新的特征变量，即主成分 z_i（$i=1\sim p$，$p<m$）。主成分是原变量 x_i（$i=1\sim m$）的线形组合且具有正交特性：

$$Z = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} \tag{20-2}$$

其中 z_1，z_2，…，z_p 按方差比例依次称为原变量的第一、第二、…和第 p 主成分。实际操作时，往往挑选几个方差比例最大的主成分，这样既可减少变量的数目，又

抓住了主要矛盾。

可以看出，主成分分析的数学实质是：寻找以取样点为坐标轴，以变量为矢量的 m 维空间中椭球体的主轴，主轴即为变量之间的相似系数 r_{ij}（i，$j=1\sim m$）矩阵中 p 个较大特征值所对应的特征向量。通常，可以用雅可比（Jacobi）法计算特征值和特征向量。

与此类似，还有一种主因子分析技术，它是以变量作为坐标轴，以取样点作为矢量，通过取样点之间的相似系数建立相关矩阵，来研究取样点之间的亲疏关系，进而找出代表性的取样点。

20.1.2 关键变量分析算法

关键变量分析则是利用变量之间的相似系数建立相关矩阵，通过用户确定的阈值、从数据库变量集中找出一定数量的关联独立变量，进而消除其他冗余变量。

设有 n 个采样点，每个采样点有 m 个变量。变量之间的关系可以用相关系数 $r_{ij}(i,j=1\sim m)$表示，r_{ij}为变量 x_i、x_j 之间数据标准差标准化后的夹角的余弦($0\leqslant r_{ij}\leqslant 1$)：

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})(x_{kj}-\overline{x_j})}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})^2\sum_{k=1}^{n}(x_{kj}-\overline{x_j})^2}} \tag{20-3}$$

显然，$|r_{ij}|$越接近于1，说明变量之间关系越密切；$|r_{ij}|$越接近于0，则变量之间关系越疏远。选定某一阈值 t，比如 $t=0.01$，0.04，0.09，0.16，0.25……等，就可以从相关矩阵中将关系疏远的变量逐个挑选出来。以表 20-1 所示的观测变量相关矩阵为例，关键变量的分析过程为：

(1) 将相关矩阵中对角线之下（$j>i$）的所有元素 r_{ij}的值取平方 r_{ij}^2；

(2) 在新的平方矩阵中，选取 r_{ij}^2的最小值所对应的两个变量 x_i 和 x_j 为两个关键变量；

(3) 将其他所有与变量 x_i 和 x_j 有联系，且 $r_{ij}^2>t$ 的变量均从变量表中删除；

(4) 将剩余变量中 x_i 和 x_j 有联系，且平方最小的相关系数所对应的两个变量选为两个关键变量；

(5) 重复第3、4步；直到全部变量均经过处理，或者关键变量个数已满足要求为止。

若预设 $t=0.25$，则按本算法由表 20-1 得到的前 8 个关键变量依次为 x_8、x_{14}、x_7、x_1、x_4、x_5、x_{13}、x_{18}。

20.1.3 变量聚类分析算法

变量聚类分析是将一组采样点或变量，按其亲疏程度进行分类。采样点或变量的相似性可以用欧几里德距离（Euclidean distance）Ed_{ij}、马氏距离（Mahalonobis distance）Md_{ij}、切比雪夫距离、兰氏距离或绝对距离等来进行度量。此处重点介绍基于欧几里

表 20-1　采样点的欧几里德距离矩阵 $ed(0)$

（据黄杏元等，2001）

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1																			
2	0.4197	1																		
3	-0.0555	0.0215	1																	
4	0.0146	0.2443	0.5520	1																
5	-0.1286	-0.1258	0.6521	0.0017	1															
6	0.9148	0.4794	-0.0013	0.0897	-0.1336	1														
7	-0.1272	-0.2901	0.0923	-0.1628	0.0912	-0.1295	1													
8	-0.0216	-0.2681	-0.4527	-0.4438	-0.1182	-0.1131	-0.3763	1												
9	-0.0955	0.5102	-0.0846	-0.3741	-0.1469	-0.1945	0.7495	0.0536	1											
10	-0.1983	-0.4163	0.3712	0.3499	0.0886	0.1152	0.0924	-0.3346	0.1729	1										
11	0.1419	-0.1401	0.2758	0.1308	0.1235	-0.1341	0.7621	-0.6021	0.3938	0.2742	1									
12	-0.0505	-0.0278	0.2721	0.4630	0.0872	-0.0007	-0.4045	-0.2361	-0.4099	0.3686	0.0673	1								
13	0.4632	0.5854	0.1813	0.0119	0.0425	0.4931	0.3259	0.1443	0.3663	0.2584	0.1973	0.1661	1							
14	-0.1989	0.0288	-0.5728	-0.1747	-0.3862	-0.2744	0.0075	-0.0004	-0.0362	-0.2117	0.0247	-0.1727	-0.4110	1						
15	-0.0900	0.2699	0.2647	0.1417	0.2550	-0.0119	0.0613	-0.2873	-0.1113	0.0622	0.1865	0.2015	0.2092	-0.2986	1					
16	0.1587	-0.4601	-0.4146	-0.1624	-0.3781	0.1538	0.1325	0.2865	0.1152	0.1069	0.1982	0.0814	0.1069	0.1082	0.1984	1				
17	0.1612	0.4005	0.3855	0.2108	0.3742	0.1481	-0.1162	-0.3126	-0.0262	-0.0164	0.1602	0.0853	0.0309	0.1270	0.1187	-0.9357	1			
18	-0.0201	-0.3388	-0.4792	0.3679	0.2352	0.0902	0.1987	0.2145	0.3245	-0.0492	0.0280	0.2987	-0.1219	0.1823	0.2686	0.5614	0.5396	1		
19	0.2291	0.6400	0.5935	0.4100	0.2019	0.3345	0.0702	-0.5273	0.3254	0.0692	0.2189	0.0437	0.4329	-0.3393	0.4440	-0.5603	0.4703	-0.5749	1	
20	0.2724	0.5739	0.6306	0.4543	0.1860	0.3827	0.0435	-0.4884	-0.3265	0.0099	0.2436	0.1031	0.4599	-0.4296	0.4151	-0.4744	0.3926	-0.5286	0.9535	1

德距离和马氏距离的变量聚类与采样点聚类算法

1. 变量聚类分析算法

设任意两个变量 x_i、x_j（$i, j=1\sim m$）在 n 维采样空间的相似性可以用欧几里德距离 Ed_{ij} 或马氏距离 Md_{ij} 来度量：

$$Ed_{ij}=\sqrt{\sum_{k=1}^{n}(x_{ik}-x_{jk})^2} \tag{20-4}$$

式中：k 为采样点的编号（$k=1\sim n$）；x_{ik}，x_{jk} 为变量 x_i、x_j 在第 k 号采样点的数据值。

$$Md_{ij}=(X_i-X_j)'\sum{}^{-1}(X_i-X_j) \tag{20-5}$$

式中：X_i，X_j 为变量 x_i、x_j 对应 n 个采样点的数据向量；$\sum^{-1}$ 为逆协方差矩阵。

距离 Ed_{ij} 或 Md_{ij} 越小，说明两个变量的相似性越大。

2. 采样点聚类分析算法

设任意两个采样点 i，j（$i, j=1\sim n$）在 m 维变量空间的相似性可以用欧几里德距离 Ed_{ij} 或马氏距离 Md_{ij} 来度量：

$$Ed_{ij}=\sqrt{\sum_{k=1}^{m}(x_{ik}-x_{jk})^2} \tag{20-6}$$

式中：k 为变量的编号（$k=1\sim m$）；x_{ik}，x_{jk} 为采样点 i，j 的第 k 号变量的数据值。

$$Md_{ij}=(X_i-X_j)'\sum{}^{-1}(X_i-X_j) \tag{20-7}$$

式中：X_i，X_j 为采样点 i，j 对应 m 个变量的数据向量；$\sum^{-1}$ 为逆协方差矩阵。

距离 Ed_{ij} 或 Md_{ij} 越小，说明两个采样点的相似性越大。

以基于欧几里德距离 Ed_{ij} 的采样点聚类分析为例，算法的基本步骤为：

(1) 计算采样点之间的欧几里德距离 Ed_{ij}，形成距离矩阵 ED (0)；

(2) 选择 ED (0) 中的非对角最小元素，设为 Ed_{pq}，则将 p，q 并为一类，记为 $G_r=\{G_p, G_q\}$；

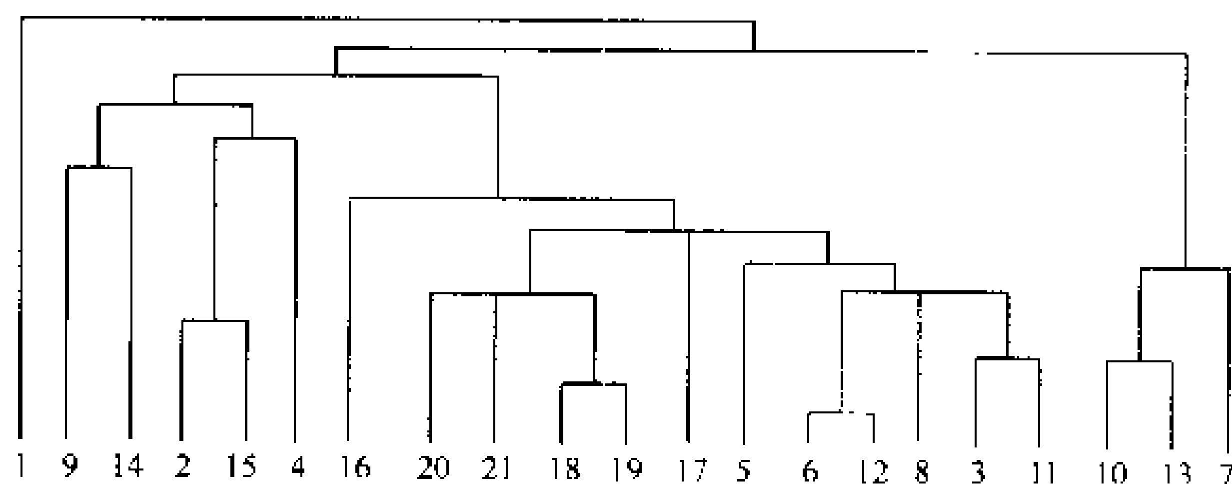


图 20-1　采样点聚类结果

（据黄杏元等，2001）

表 20-2 采样点的欧几里德距离矩阵 *ed*(0)

(据黄杏元等,2001)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0																				
2	1.82	0																			
3	3.22	1.35	0																		
4	2.37	0.51	0.91	0																	
5	3.00	1.18	0.41	0.73	0																
6	3.58	1.66	0.33	1.23	0.72	0															
7	4.23	2.49	1.18	1.98	1.31	0.89	0														
8	3.56	1.82	0.51	1.31	0.74	0.32	0.66	0													
9	1.64	0.52	1.95	1.05	1.76	2.28	2.65	2.06	0												
10	4.04	2.16	0.83	1.69	1.22	0.54	0.35	0.78	2.74	0											
11	3.11	1.27	0.20	0.78	0.33	0.51	1.32	0.65	1.81	1.01	0										
12	3.56	1.64	0.31	1.21	0.75	0.08	0.97	0.30	2.26	0.60	0.45	0									
13	4.18	2.34	1.03	1.85	1.40	0.74	0.33	0.76	2.72	0.20	1.17	0.82	0								
14	2.06	1.10	2.43	1.53	2.08	2.76	2.93	2.24	0.48	2.82	2.29	2.74	3.20	0							
15	1.79	0.27	1.60	0.70	1.29	1.93	2.44	1.79	0.47	2.39	1.46	1.91	2.45	0.83	0						
16	3.95	2.17	0.86	1.66	1.17	0.57	0.52	0.43	2.33	0.67	1.00	1.70	0.65	2.55	2.22	0					
17	2.89	1.15	0.78	0.64	0.39	1.11	1.60	1.05	1.65	1.57	0.64	1.09	1.21	1.97	1.18	1.48	0				
18	2.94	1.20	0.91	0.69	0.66	1.24	1.51	0.95	1.36	1.70	0.71	1.22	1.68	1.52	1.71	1.53	0.44	0			
19	2.84	1.10	0.83	0.69	0.54	1.16	1.39	0.84	1.26	1.62	0.68	1.14	1.60	1.66	1.07	1.15	0.43	0.14	0		
20	3.45	1.67	0.38	1.16	0.67	0.72	1.02	0.47	1.83	1.17	0.49	0.69	1.15	2.05	1.72	0.50	0.93	0.53	0.65	0	
21	3.22	1.38	0.59	0.89	0.46	0.92	1.31	0.76	1.58	1.33	0.45	0.90	1.38	1.83	1.49	0.81	0.75	0.32	0.42	0.31	0

(3) 计算新类及其他类的距离：将 ED (0) 中的第 p、q 行和第 p、q 列删除，并在第 p 行 q 列的位置上记上 D_{rk} ($k=1, 2, \cdots, m$；$k \neq p$、q)，形成新矩阵 ED (1)；

(4) 对新矩阵 $ED(1)$ 重复关于 $ED(0)$ 的步骤；得到新矩阵 $ED(2)$、$ED(3)$…直到所有采样点均得到归类为止。

需要注意的是，如果某一步的 ED (i) 中的最小元素不止 1 个，则其所对应的采样点均归为一类。

以表 20-2 所示的采样点欧几里德距离矩阵 ED (0) 为例，经过 14 次矩阵运算和聚类，最后得到的归类结果如图 20-1 所示。

§20.2 空间分类统计算法

空间分类统计是基于地图表达，采用与变量聚类分析相类似的方法来产生新的综合性或简洁性专题地图的过程。空间分类统计包括空间聚类、空间聚合和判别分析 3 类。

20.2.1 空间聚类分析算法

空间聚类分析的基本思想是：在栅格地图的基础上，经过对两个或两个以上变量的逻辑运算，将符合某种预设聚类条件的新栅格做地图输出，而不符合聚类条件的则区域空白。其算法表达式为：

$$Ce(U) = \{(A,P) \in U \mid (A,P) \text{ 满足 } e\} \tag{20-8}$$

式中：U 为变量的栅格数据集；A 为变量集；P 为游程标识；e 为聚类条件集。

以基于二元变量 (A_1, A_2) 和三元变量 (A_1, A_2, A_3) 的布尔逻辑运算为例，其逻辑表达如图 20-2 所示。布尔逻辑运算遵循以下基本定律：

(1) 交换律：$A \cup B = B \cup A$；$A \cap B = B \cap A$

(2) 分配律：$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(3) 结合律：$(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$

$(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$

(4) Demogan 定律：$A - (B \cup C) = (A - B) \cap (A - C)$

$A - (B \cap C) = (A - B) \cup (A - C)$

以矿山 GIS 为例，有井田煤层分布的栅格图若干，其中 A_1、A_2、A_3 分别代表煤层厚度、煤的发热量和煤层埋深。现要进行开采决策分析，给定聚类条件 $e = \{A_1 \geq 5\text{m and } A_2 \geq 7000\text{cal}\}$，则要求将井田内煤层厚度大于 5m、煤的发热量大于 7000 大卡的优质煤层分布区域圈定出来；若给定聚类条件 $e = \{A_1 \geq 5\text{m and } A_2 \geq 7000 \text{ cal and } A_3 \leq 500 \text{ m}\}$，则进一步要求将井田内煤层厚度大于 5m、煤的发热量大于 7000 大卡且埋深不大于 500 m 的浅部优质分布区域圈定出来，供开采设计和决策参考。

20.2.2 空间聚合分析算法

空间聚合分析的基本思想是：根据地图的空间分辨率或属性分类表进行数据类别合

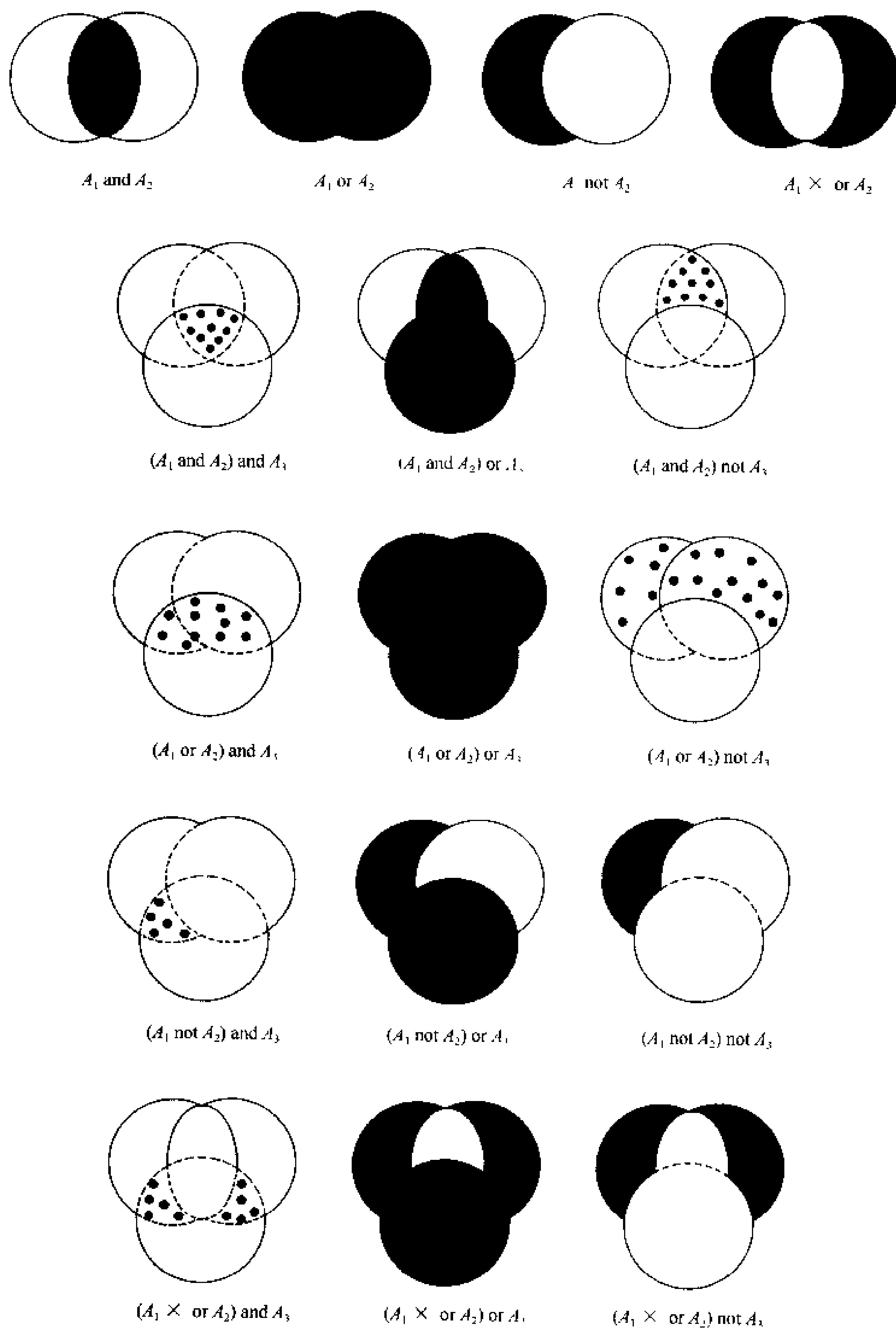


图 20-2 基于二元变量和三元变量的空间聚类的逻辑运算示例

并或转换，以实现空间地域的兼并。空间聚合分析的结果是将复杂的属性类别转化为简单的属性类别，并以更小比例尺输出专题地图。空间聚合包括分类等级的粗化、数据容差的扩大和细部合并等过程。因而，在某种意义上说，空间聚合类似于地图综合，也可以说是地图综合技术在 GIS 空间统计分析中的应用扩展。

图 20-3（a）为一幅按人口密度 100 人/km² 为间距的 1:5000 的乡村级人口统计分区图，若按人口密度 200 人/km² 为间距，则转化为如图 20-3（b）所示的 1:25000 的县市级人口统计分区图。

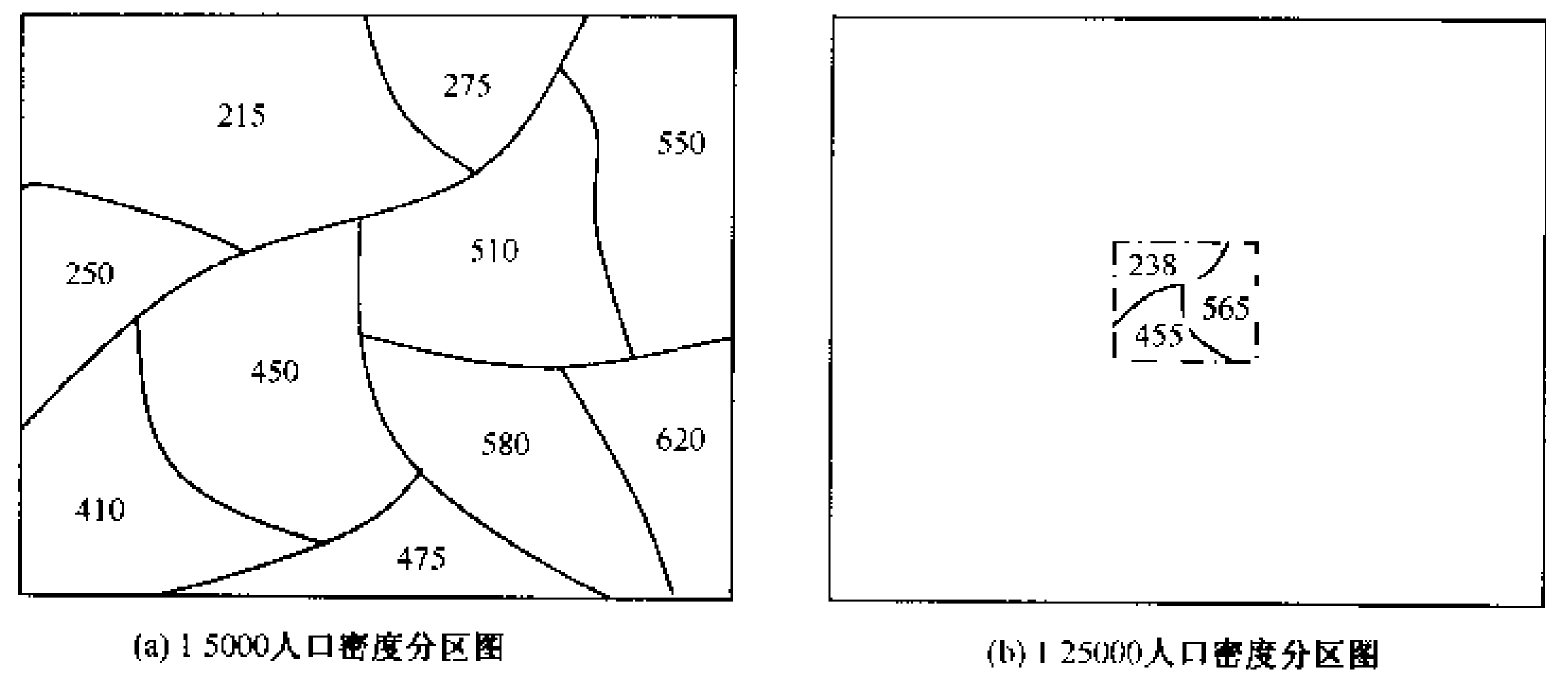


图 20-3　空间聚合分析示例

20.2.3　判别因子分析算法

判别分析区别于聚类分析之处是：根据预先确定的等级序列因子标准和判别临界值，将待分析的对象进行分析判别，并将其划归到序列中的合理位置。其具体应用领域包括水土流失评价、土地适宜性评价、矿体可采性评价、环境容量评价等。

设评判对象集的属性要素（变量）为 x_i（$i=1\sim m$），各要素的评价权（综合反映属性要素的作用力和贡献率）为 p_i（$i=1\sim m$），则可以构造一个线形判别函数：

$$Y=\sum_{i=1}^{m}a_i x_i \tag{20-9}$$

按判别因子 Y 的大小排队，则可以实现对评判对象集的简单分类。但是，为使各分类之间的界限尽可能分明，而且各分类对象之间尽可能接近，应设定一个分类临界值。设有 A、B 两类，其类中对象数目分别为 n_1 和 n_2，类中各对象的判别因子值分别为 $Y_i(A)(i=1\sim n_1)$、$Y_i(B)(i=1\sim n_2)$，两类的判别因子平均值分别为 $\bar{Y}(A)$、$\bar{Y}(B)$，应满足以下比值条件：

$$I=\frac{[\bar{Y}(A)-\bar{Y}(B)]^2}{\sum_{i=1}^{n_1}[Y_i(A)-\bar{Y}(A)]^2+\sum_{i=1}^{n_2}[Y_i(B)-\bar{Y}(B)]^2}=T \tag{20-10}$$

式中，T 为预先设定的分类临界值。

由此可见，判别因子分析的分类过程为：① 首先构建判别函数，确定属性要素及

其评价权重；② 计算各对象的判别因子并排序，按排序结果进行简单的初始分类；③ 检查比值是否大于分类临界值：若非，则调整初始分类后再检查；若是，则动态调整分类结果，直到满意为止。

§20.3 层次分析算法

GIS 中层次分析法（hierarchical analysis procedure，AHP）实质是一种基于地理区域的、模仿人的思维过程的、定性定量分析相结合的系统分析方法。该法把人的地理思维过程层次化和数量化，并用数学方法进行描述，进而为各类与地理区域相关的分析、决策、预报和控制提供定量依据。通常，层次分析中的最高层为目标层，该层只有一个目标元素，当然，该目标也可以分解为若干个子目标；中间层为准则层，是为实现目标而采取的各类策略、约束、准则或因素的综合影响模式，也可以按分类等级将其分为若干子层；最低层为指标层，描述影响目标实现的各有关因素。

研究与地理空间相关的问题往往涉及大量相互关联、相互制约的复杂因素，而且各因素对问题的分析具有不同的重要性。AHP 法的核心是：将相互关联的要素按隶属关系分为若干层次，并根据专家知识对不同层次的相对重要性给出定量指标，然后利用数学方法综合专家意见给出各层次要素的相对重要性权值，形成判断矩阵，以此作为综合分析和方案比较及形成决策的基础。

1. 判断矩阵构建及其一致性修正

设某一目标问题 Y 有 m 个影响因素 x_i（$i=1\sim m$），其中每两个因素 x_i 和 x_j（$i, j=1\sim m$）对 Y 的影响权之比为 a_{ij}（$i, j=1\sim m$），满足 $a_{ij}>0$，$a_{ii}=1$ 且 $a_{ji}=1/a_{ij}$。由 $(a_{ij})_{m\times m}$ 所组成的矩阵 A 称为正互反矩阵

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \tag{20-11}$$

如果该矩阵满足

$$a_{ij}\times a_{jk}=a_{ik}\cdots(i,j,k=1\sim m) \tag{20-12}$$

那么，A 具有一致性。数学上可以证明：m 阶一致矩阵 A 的最大特征根 $\lambda_{\max}=m$。

实际上，在构建判断矩阵的过程中，由于要素的众多及其关联的复杂性，以及专家认识的主观片面性或逻辑非一致性，往往导致所构造的矩阵不满足一致性要求。此时，需要进行一致性判断和修正。

当所构造的判断矩阵非一致时，其特征根 $\lambda_{\max}>m$。建立以下一致性检验指标：

$$CI=\frac{\lambda_{\max}-m}{m-1} \tag{20-13}$$

$CI=0$，说明所构造的判断矩阵是一致的；CI 越大，说明非一致性越强。

Saaty T. L.（1980）采用随机抽样的方法，得出不同阶数（m）矩阵的平均随机

性指标 RI，如表 20-3 所示。

表 20-3　平均随机性一致指标 RI

m	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

令

$$CR = \frac{CI}{RI} \tag{20-14}$$

CR 称为随机一次性比率。当 $CR<0.1$ 时，认为所构造的判断矩阵是可以接受的；否则，应修改判断矩阵中各元素的值，使其满足可接受的要求。

2．要素影响权比的确定

确定要素影响权比的方法有多种，如根据专家所确定的各要素的权重 w_i（$i=1\sim m$）来计算两因素 x_i 和 x_j 对 Y 的影响权的比值 a_{ij}；或采用 Saaty T. L. 的办法。Saaty T. L.（1980）引用数字 1～9 及其倒数作为标度来确定 a_{ij}，如表 20-4 所示。

表 20-4　两因素的影响权之比

x_i/x_j	相等	之间	较强	之间	强	之间	很强	之间	极强
a_{ij}	1	2	3	4	5	6	7	8	9

以矿产资源开发为例，假设矿区建设投资拟考虑的影响因素有矿产储量、矿产品位、交通运输条件、土地开发容量、水资源条件、地质复杂程度和环境承受能力等 7 项因素。根据专家意见和综合知识，考虑到影响因素两两之间对矿区建设投资的比重，所建立的判断矩阵如式（20-15）所示。可以查证，该判断矩阵虽为正互反矩阵，但并不满足一致性要求，需要进行一致性修正。

$$A=\begin{bmatrix} 1 & 2 & 3 & 5 & 2 & 3 & 2 \\ 1/2 & 1 & 5 & 4 & 1 & 5 & 3 \\ 1/3 & 1/5 & 1 & 2 & 1/4 & 1/3 & 1/2 \\ 1/5 & 1/4 & 1/2 & 1 & 1/3 & 1/2 & 1 \\ 1/2 & 1 & 4 & 3 & 1 & 3 & 2 \\ 1/3 & 1/5 & 3 & 2 & 1/3 & 1 & 1/2 \\ 1/2 & 1/3 & 2 & 1 & 1/2 & 2 & 1 \end{bmatrix} \tag{20-15}$$

3．动态判断与残缺判断

由于实际地理相关现象的复杂性和动态性，必然导致其影响因素的影响权之比也随时间发生变化。此时，需要构造一个基于时间函数的判断矩阵，该矩阵即称为动态判断矩阵。该矩阵中的各元素 a_{ij}（t）的函数形式可以参照下表选择。

表 20-5 两因素的影响权之比

(据边馥苓，1996)

$a_{ij}(t)$ 的函数	含 义	注 释
a	与 t 无关	为整数，$1 \leqslant a \leqslant 9$
$a_1t + a_2$	为 t 的线形函数；升或降到某一点之后成为常数；其反商是双曲函数	某一因素比另一因素的重要性稳定增长
$a_1\lg(1+t) + a_2$	对数上升（降）到某一点之后成为常数	先迅速增加（降低），然后缓慢增加（降低）
$a_1e^{a_2t} + a_3$	指数增加到最大（a_2 为负数时衰减），然后保持常数	先缓慢增加（降低），然后迅速增加（降低）
$a_1t^2 + a_2t + a_3$	一个有极大极小值的抛物线（$a_1>0$ 时有极小值；$a_1<0$ 时有极大值）；然后是常数	增加（减少）到极大值（极小值），然后下降（增加）
$a_1t\sin(t+a_2) + a_3$	正弦振荡	振幅周期性增大或减少
剧变	不连续	无规律剧烈变化

表中的参数是根据经验数据经曲线拟合得到的。

某些情况下，比如当信息不足和知识不够，可能对判断矩阵中的某些项把握不准而无法确定，造成该项元素残缺。此时的判断矩阵即为残缺矩阵。如果一个残缺矩阵的任一残缺元素均可以通过矩阵中的已知信息间接获得，则该残缺矩阵是可接受的；否则，不可接受。

可接受残缺矩阵的最大特征根和权重特征向量 $W = (w_i)_{m\times 1}$ 的算法如下：

设 $A = (a_{ij})_{m\times m}$ 是残缺判断矩阵，θ_i 为其中的残缺元素。构造一个辅助矩阵 $C = (c_{ij})_{m\times m}$ 如下：

$$C = \begin{cases} a_{ij}, (a_{ij} \neq \theta, i \neq j) \\ 1, (i = j) \\ w_i / w_j, (a_{ij} = \theta, i \neq j) \end{cases} \tag{20-16}$$

求辅助矩阵 C 的最大特征根 $\lambda_{\max}^{C}$，使得

$$CW = \lambda_{\max}^{C} W \tag{20-17}$$

由于辅助矩阵 C 与等价矩阵 $\bar{C}$ 有相同的特征根及其对应的特征向量，所以可以通过求等价矩阵 $\bar{C}$ 的特征根来确定辅助矩阵 C 的最大特征根 $\lambda_{\max}^{C}$：

$$\bar{C} = \begin{cases} a_{ij}, (a_{ij} \neq \theta, i \neq j) \\ 0, (a_{ij} = \theta) \\ n_i + 1, (i = j, n_i \text{ 为第 } i \text{ 行中残缺元素的个数}) \end{cases} \tag{20-18}$$

其最大特征根 $\lambda_{\max}^{\bar{C}}$ 满足

$$\bar{C}W = \lambda_{\max}^{C} W \tag{20-19}$$

例：设有可接受的残缺判断矩阵

$$A = \begin{bmatrix} 1 & 2 & \theta_1 \\ 1/2 & 1 & 2 \\ \theta_2 & 1/2 & 1 \end{bmatrix}$$

构造其辅助矩阵

$$C = \begin{bmatrix} 1 & 2 & w_1/w_3 \\ 1/2 & 1 & 2 \\ w_3/w_1 & 1/2 & 1 \end{bmatrix}$$

确定辅助矩阵的等价矩阵

$$\bar{C} = \begin{bmatrix} 2 & 2 & 0 \\ 1/2 & 1 & 2 \\ 0 & 1/2 & 2 \end{bmatrix}$$

解得 $\bar{C}$ 的特征根和权重特征向量分别为

$$\lambda_{\max}^{C} = 3, \quad W = \begin{bmatrix} 0.5714 \\ 0.2857 \\ 0.1429 \end{bmatrix}$$

可以用下式对 $\bar{C}$ 进行一致性检验：

$$CI = \frac{\lambda_{\max} - n}{(m-1) - \dfrac{\sum_{i=1}^{m} n_i}{m}} \tag{20-20}$$

当 $CR<0.1$ 时，$\bar{C}$ 具有满意的一致性。

参 考 文 献

边馥苓. 1996. 地理信息系统原理和方法. 北京：测绘出版社. 140～148

龚健雅. 2001. 地理信息系统基础. 北京：科学出版社. 272～278

黄杏元，马劲松，汤勤. 2001. 地理信息系统概论. 北京：高等教育出版社. 183～193

李庆扬，王能超，易大义. 1982. 数值分析. 武汉：华中理工大学出版社. 61～117

沈永欢，梁在中，许履瑚，蔡倩倩. 2002. 数学手册. 北京：科学出版社

Saaty T L, 1980. The Analytic Hierarchy Process. McGraw Hill International Book Company

第二十一章　空间分析算法

空间分析是GIS的一项十分重要的任务和最具特色的功能，是基于地理目标的位置和形态特征的空间数据分析技术，其目的是提取和发现隐含的空间信息或规律，是空间数据挖掘和知识发现的基本方法之一（见第23章）。空间分析不是简单地通过“检索”、“查询”或“统计”从地理数据库中提取时空信息，而是利用各种空间分析模型及空间操作对地理数据库中的空间数据进行深加工，进而产生新的知识。空间分析是空间数据的空间特性与非空间特性的联合分析，亦即拓扑与属性数据的联合分析，主要涉及基于网络的路径分析、资源分配分析，以及缓冲区分析和叠置分析4类。

§21.1　路径分析算法

图论是研究事物及其之间关系的科学，任何一个能用二元关系描述的系统，且都可以用图进行建模。网络分析是基于图论和运筹学的技术，它通过研究网络状态，来模拟分析和研究资源在网络上的流动、分配情况，进而实现对网络结构及资源分配等的优化。通常，网络分析包括路径分析和资源分配分析等两大类，涉及最短路径、最优路径、网络定位、网络分配、结点或弧段的游历、最小连通树、最大（小）物流等问题。

21.1.1　地理网络的基本概念

地理网络是地理空间中的一类具体的网络系统，图论中的一些基本概念在这里也同样适用；除此之外，还有地理网络自身的一些特殊概念。

1. 一般网络的基本要素

(1) 结点/结点集：网络中任意两条线段的交点为结点 v_i。网络系统（G）中所有结点的集合称为结点集 $V(G)=\{v_1, v_2, \cdots v_n\}=[v_1\ \ v_2\cdots v_n]^T$。

(2) 边/边集（若边有方向，则为弧/弧集，用 a_i 和 A 表示；为叙述方便，以下简称边/边集）：网络中任意一条线段为边 e_i。网络系统（G）中所有边的集合称为边集 $E(G)=\{e_1, e_2, \cdots e_n\}=[e_1\ \ e_2\cdots e_n]^T$。边以点集中的某两个点为起点和终点，即 $e_{ij}=v_iv_j$，故边集也可以表示为：$E(G)=\{(v_i, v_j)/v_i\in V, v_j\in V\}$。若边的两个端点重合，该边称为环；若两条边的端点是同一对结点，则这两条边称为重边或重弧。

(3) 图：图是一个非空的有限结点与有限边的集合，表示为 $G(V,E)$。不考虑边的方向的图称为基础图，记为 $G(V,E)$；考虑边的方向的图称为有向图，记为 $D(V,A)$；既没有环也没有重弧的有向图称为简单有向图。

(4) 网络：给定有向图 $D(V,A)$，如果对图中的每一条弧 a_i 和结点赋予一个实数

权重 $w(a_i)$ 或 $w(v_i)$，则称为赋权有向图，即网络，记为 $D=(V, A, W)$。$W = W(D) = \{w_1, w_2, \cdots w_m\} = [w_1 \quad w_2 \cdots w_m]^T$ 称为D 的权函数或权矩阵；只给网络中的弧赋权或只给网络中的结点赋权的网络，分别称为弧权网络或点权网络。

（5）流：指网络 D 中任意一条弧 $a_{ij}=(v_i, v_j)$的物流量，记为 $f(a_{ij})=f_{ij}$，

$$f_{ij} = \sum_{(v_i, v_j) \in A} f_{ij} - \sum_{(v_j, v_i) \in A} f_{ji} \tag{21-1}$$

网络 D 的总流量满足

$$\sum_{(v_i, v_j) \in A} f_{ij} - \sum_{(v_j, v_i) \in A} f_{ji} = 0 \cdots (v_i \in V \setminus \{v_s, v_t\}) \tag{21-2}$$

2．地理网络的特殊要素

（1）站点：网络路线中物流装、卸的位置（不一定在结点处），如公共交通网络中的公共汽车站、邮政网络中的邮件投放点等；

（2）中心：网络图中具有接收或分发物流能力的核心结点，如水系网络中的水库，交通体系中的学校和小区等；

（3）障碍点：限制物流通过的结点，如河流的水闸，通风系统的风门等；

（4）转弯点：网络系统中物流方向发生改变的结点。若某转弯点是 n 条弧的公共结点，则该转弯点可能产生的转弯方向最多有 n^2。实际上，转弯点往往有方向控制，如禁左、禁右、禁掉头等，所以其可能的转弯方向远远少于 n^2。

（5）段（section）：段是一条弧或弧的某一部分，段同样有起始位置和终止位置，可以通过段的长度与其所在该弧段的长度的百分比来度量。

（6）路径（route）：路径是定义了属性的有序弧的集合（至少应包含某一条弧或其中的一部分），表示一个线形特征，如学院路口—成府路东口路段、黄村—李庄路段，107 桩—125 桩路段等。路径是地理网络中具有较完整意义的特征子类，可以与各种事件直接关联。结点不重复的路径称为简单路径。

（7）路径系统（route-system）：是路径和段的集合，通常用来管理具有相同属性的多个线形特征。如一个城市的所有公交行车路线可以看成一个路径系统。一个路径系统要使用统一的度量标准（如距离、时间等）进行数据管理与分析。

3．地理网络要素的属性

网络要素的属性除了一般 GIS 所要求的名称、关联要素、方向、拓扑关系等空间属性之外，还有一些特殊的非空间属性，如：

（1）阻强（impedance）：指物流在网络中运移的阻力大小，如所花时间、费用等。阻强一般与弧的长度、弧的方向、弧的属性及结点类型等有关。转弯点的阻强描述物流方向在结点处发生改变的阻力大小，若有禁左控制，表示物流在该结点的往左运动的阻力为无穷大或为负值。为了网络分析需要，一般要求不同类型的阻强要统一量纲。

（2）资源需求量（demand）：指网络系统中具体的线路、弧段、结点所能收集的或可以提供给某一中心的资源量。如供水网络中水管的供水量，城市交通网络中沿某条街道的流动人口；货运站的货量等。

(3) 资源容量：指网络中心为满足各弧段的要求所能提供或容纳的资源总量，也指从其他中心流向该中心或从该中心流向其他中心的资源总量。如水库的容量、货运总站的仓储能力等。

停靠点仅在选择最佳路线时使用，其属性有资源需求量，正值表示装载，负值表示卸载。而中心点仅在寻求网络最佳状态时使用，其属性包括资源最大容量、服务范围（从中心至各可能路径的最大距离）和服务延迟数（在其他服务中心达到某项临界值时开始起动服务）。

(4) 事件 (event)：路径系统中的某一路径的分段属性。这些属性由用户定义，并用路径的度量来表示。事件分为点事件（与一个位置对应，用一个度量表示）、线事件（用两个度量表示一个区段）和连续事件（用一个度量表示一个区段的开始和下一个区段的开始）3类。

21.1.2 最短路径算法

路径分析是GIS的基本功能之一，其核心是最优路径的求解。路径分析中有很多著名的应用，如边最优游历方案和结点最优游历方案。边最优游历方案的实质是给定一个边集合和一个结点，使之由指定结点出发至少经过每条边一次而回到起始结点，图论称之为中国邮递员问题。结点最优游历方案的实质是给定一个起始结点、一个终止结点和若干中间结点，求解最优路径，使得由起始出发不重复遍历所有中间结点并到达终点，也称推销员问题。

所谓最短路径，是指在网络 D (V, A) 中，找出从起点 v_s 出发到终点 v_e 的累计行程最短的路径。最短路径的求解算法一般分为两大类，即单源点间的最短路径 (single source shortest path) 和所有点对之间的最短路径 (all pairs shortest path)，前者是求网络系统中某一点到其他点的最短路径；而后者是在整个运算过程中，求得所有点对之间的最短距离。

1. 单源点的最短路径

Dijkstra E. W. 曾提出了一个按路径长度递增的次序产生最短路径的算法。该算法的基本思路为：假设网络 D (V, A) 每个结点 v_i 都有一对标号 (d_i, p_i)，其中 d_i 是从起点 v_s 到该点的最短路径的长度，p_i 则是从起点 v_s 到该点的最短路径中的 v_i 的前一结点。算法的基本过程为：

第一步：初始化

起点 s 设置为：$d_s=0$，p_s 为空；

其他所有点 i：$d_i=\infty$，记 $p_i=?$（未知）；

标记点 $k=s$（s 为起点的编号）。

第二步：距离计算

计算从所标记的点 k 到与其直接相连的所有其他未作标记的点 i 的距离 l_{ki}，并令

$$d_i = \min[d_i, d_k + l_{ki}] \tag{21-3}$$

第三步：选取下一点

从上述结点集中，选取 d_j 最小所对应的点为最短路径中的下一连接点 j，并作标记。

第四步：找到 j 的前一点

从已标记的所有点中，找到直接连接点 j 的前一点 i^*，并令

$$j = i^* \quad 作为前一点$$

第五步：标记点 j。如果所有的点均已标记，则最短路径寻找结束；否则，记 $k = j$，转第二步继续。

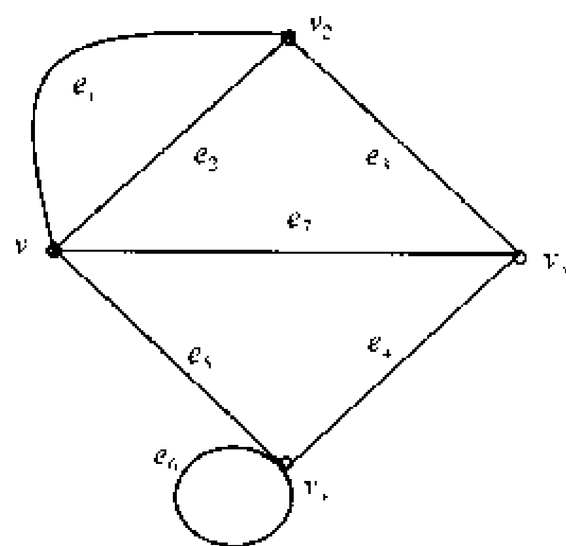
可见，该算法在求解从起点到某一终点的最短路径的过程中，还可以得到从该起点到其他各结点的最短路径。该算法的时间复杂度为 $O(n^2)$（n 为网络结点总数）。

2．多点对间的最短路径

求解网络系统中多点对乃至所有结点对之间的最短路径，可以重复执行上述的 Dijkstra 算法多次或 n 次，也可以使用 Flord 算法。但无论使用何种方法，其时间复杂度均为 $O(n^3)$。显然，这是一个非常耗时的解算过程，且运算空间的开销也非常大。

3．Dijkstra 算法的改进

Dijkstra 算法求解最短路径问题是基于网络的权矩阵，运用了关联矩阵、邻接矩阵和距离矩阵，即对于含 n 个结点的网络，运算时需要定义 $n \times n$ 阶矩阵。实际上，关联矩阵和邻接矩阵中有大量的 0 元素和∞元素，这些无效元素占用了大量的内存。以图 21-1 所示的无向简单图 $G(V, E)$ 为例，其基本的关联矩阵（结点-边关联）和邻接矩阵（结点-结点邻接）分别为：



$$M(G)_{V \times E} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 & 0 \end{bmatrix}$$

$$A(G)_{V \times V} = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

图 21-1　无向简单图示例

（据王家耀，2001）

为了提高 Dijkstra 算法的运算速度和节约对内存的开销，人们研究提出了多种改良措施，如邻接结点算法（徐立华，1989；龚洁晖、白玲，1998）、基于点-边拓扑关系的算法（徐立华，1989）、优先搜索算法（徐业昌等，1998）和采用动态数据串（距离数据串）等（丁跃民，1999）。邻接结点算法和基于点-边拓扑关系算法的出发点是要尽量减少矩阵中的 0 元素和∞元素，进而达到矩阵降维的目的，从而节约内存开销并提高运算效率。比如，可以引入一个最大极限距离 Z_{limit}，任何位于这一距离之外的点不再列

入最短路径的考虑范围。实际上，在不考虑结点数和边数的情况下，若网络中两个点的距离大于网络中相距最远的两个点的最短距离的一半以上，这两个点的距离通常可以通过其他的最短路径的组合来得到。

4. 次短路径求解算法

在某些情况下，除了要求出两个给定点之间的最短路径之外，还可能要求出这两点之间的次最短路径、第3短路径……及第k最短路径。可以在求出第1最短路径P_1之后，用枚举法求出与P_1有尽可能多公共边的次最短路径P_2。

算法的基本思路是：假定第一最短路径P_1包含了N条有向弧，每次删除其中的一条弧，即得到N个与原网络只有一弧之差的新网络。按原最短路径算法分别求解这N个新网络的最短路径，然后比较这N条最短路径，其中最短的那条即为所求的次最短路径。依次进行，可以分别求解第3短路径……及第k最短路径。

21.1.3 最佳路径算法

所谓最佳路径，是指网络两结点之间阻抗最小的路径。"阻抗最小"有多种理解，如基于单因素考虑的时间最短、费用最低、风景最好、路况最佳、过桥最少、收费站最少、经过乡村最多等，和基于多因素综合考虑的风景最好且经过乡村较多，或时间较短、路况较佳、且收费站最少等。最短路径问题是最优路径问题的一个单因素特例，即认为路径最短就是最优。

最佳路径的求解算法有几十种，如基于贪心策略的最近点接近法、最优插入法，基于启发式搜索策略的分枝算法，基于局部搜索策略的对边交换调整法，以及广泛采用的Dijkstra算法等。这里分别介绍基于最大可靠性和最大容量的最优路径（王家耀，2001）。

1. 最大可靠路径

设网络D（V，A）中的每条弧a_{ij}（v_i，v_j）的完好概率为p_{ij}，D中的任意一条路径P，其完好概率为：

$$p(P) = \prod_{a_{ij} \in F(P)} p_{ij} \tag{21-4}$$

则网络D（V，A）中所有（v_s，v_t）路径中的完好概率最大的路径为（v_s，v_t）的最大可靠路径。

利用最短路径算法也可以求解最大可靠路径。做法为：

定义网络D（V，A）中的每条弧a_{ij}（v_i，v_j）的权为

$$w_{ij} = -\ln p_{ij} \tag{21-5}$$

因为$0 \leqslant p_{ij} \leqslant 1$，所以$w_{ij} \geqslant 0$。从而可以用前述的Dijkstra算法求出关于权$w_{ij}$的最短路径。由于$\sum w_{ij} = -\ln(\prod p_{ij})$，所以，关于权$w_{ij}$的最短路径就是（$v_s$，$v_t$）的最大可靠路径，其完好概率为$\exp(-\sum w_{ij})$。

2. 最大容量路径

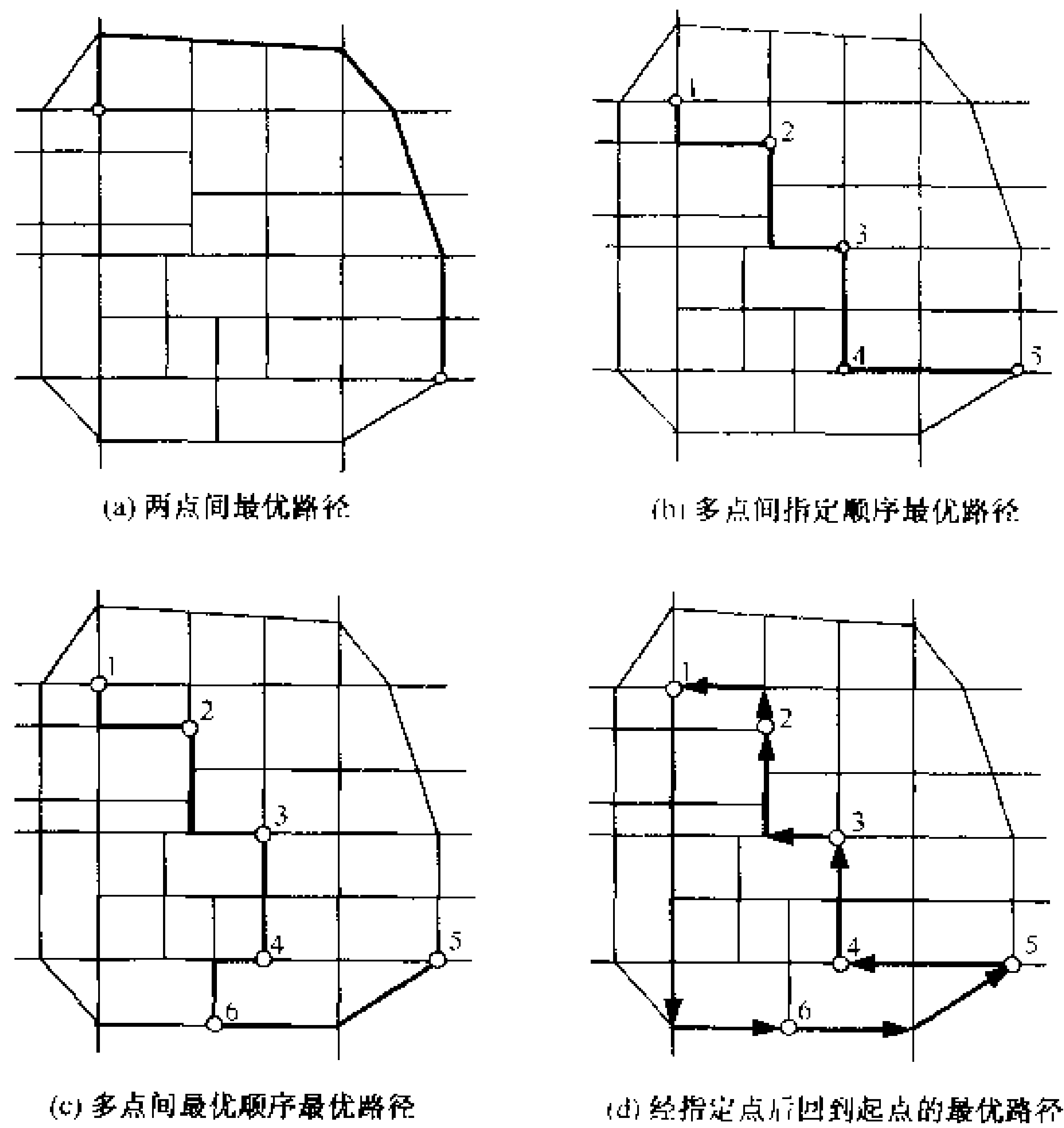
设网络 D（V, E, W）中的任意一条路径 P 的容量定义为该路径中所有弧的容量 c_{ij} 的最小值，即：

$$c(P) = \min_{(i,j) \in P} c_{ij} \tag{21-6}$$

则网络 D（V, A）中所有（v_s, v_t）路径中的容量最大的路径即为（v_s, v_t）的最大容量路径。

同样，可以将网络中每条边或弧的权值定义为通过该边或弧的时间，就可以求出时间最优路径；若定义为该弧的费用，则所求出的为费用最优路径。

最优路径的求解有多种形式（如图 21-2 所示），两点间最优路径、多点间指定顺序的最优路径、多点间最优顺序最优路径、经指定点后回到起点的最优路径等。



(a) 两点间最优路径　(b) 多点间指定顺序最优路径

(c) 多点间最优顺序最优路径　(d) 经指定点后回到起点的最优路径

图 21-2　最优路径的几种典型形式

21.1.4　最小连通树算法

最小连通树是图论中的一个基本问题，在网络规划设计中用途广泛，如在局域网建设中，如何用最少的电缆连接所有的结点。若考虑到点对之间除距离之外的其他因素，

最小连通树问题就称为最小加权连通树。

最小连通树的数学描述如下：对于一个网络 $G=(N, L)$，其中 N、L 分别为网络中的结点集和直接边集。假定 N 中任意两点 n_i、n_j 的直接边的距离或权为 $w(n_i, n_j)$。从 G 中找出一个不含圈（无回路，不重复）的子集 S，该子集是无顺序限制连接所有点的最小加权子网络：

$$W(S)|_{S\subset L} \leqslant W(T)|_{T\subset L} = \sum_{n_i, n_j \in T} w(n_i, n_j) \tag{21-7}$$

若 G 的结点总数为 n，则 S 具有 $n-1$ 条边来连接所有的点。图 21-3 所示为最小连通树的一个简单例子。

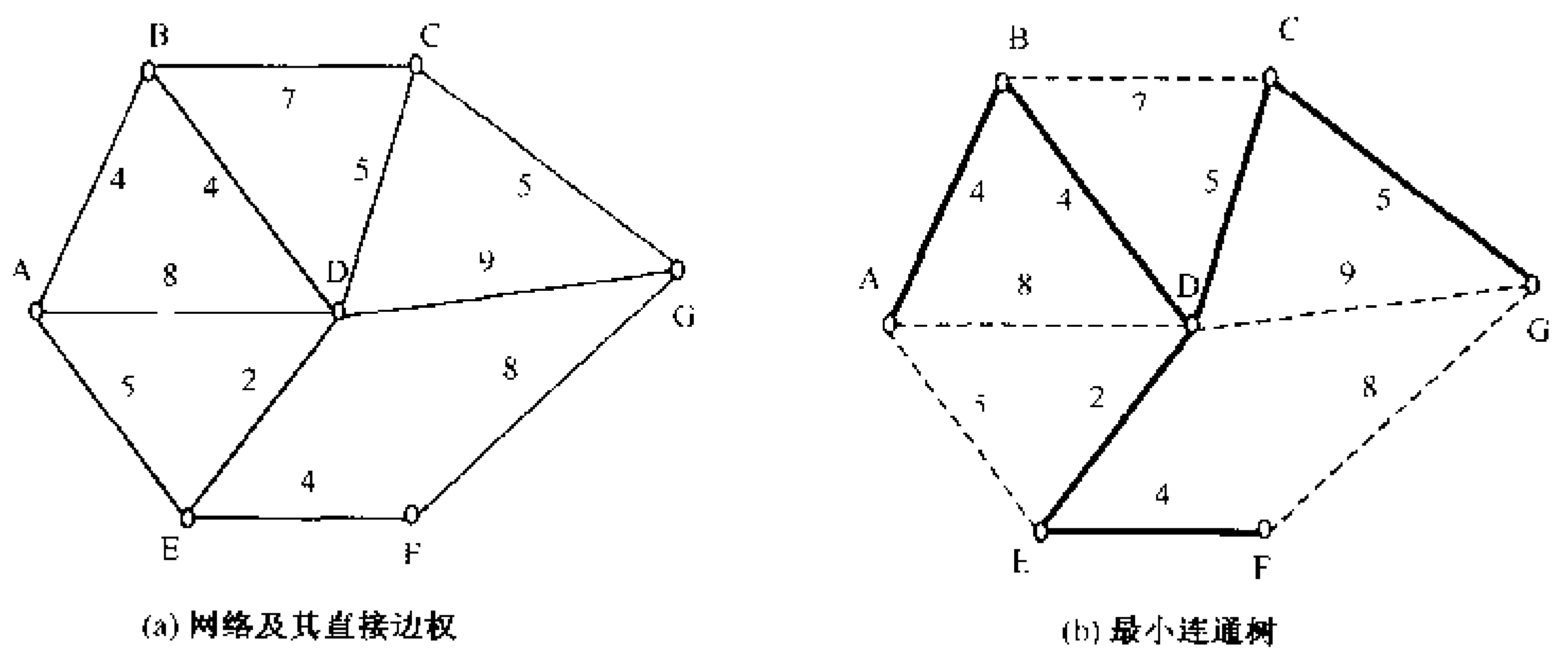


图 21-3　最小连通树的简单例子
（据丁跃民，1999）

最小连通树的求解有两种主要算法，即 Kruskal 算法和 Prim 算法。

Kruskal 算法的基本原理为：从最小权边开始，每一步从未选的直接边中选出一条最小权的边，使其与已选的边不构成圈。该算法的难点是检测所选出的边是否构成圈。

Prim 算法是对 Kruskal 算法的改进，其基本原理是构建一个动态的选中点集 M，通过对 M 的不断扩充来完成最小连通树的查找。基本步骤如下：

(1) 从点集 N 中选择一点 u，将 u 转到点集 M 中，即

$$M \leftarrow M + u, \cdots N \leftarrow N - u \tag{21-8}$$

(2) 从点集 M 与 N 之间，找出一个最小连接边 (k, l)，结点 $k \in M$，$l \in N$，即

$$(k, l) = \min[(u, v)] \mid U \in M, V \in N \tag{21-9}$$

(3) 将结点 l 选入点集 M，并记下选中的边 (k, l)，即

$$M \leftarrow M + l, \cdots N \leftarrow N - l, \cdots S \leftarrow (k, l) \tag{21-10}$$

(4) 如果点集 N 中仍有待选点，则转 (2)；否则，完成检查，输出最小连通树 S。

§21.2　资源分配算法

资源分配用来模拟地理网络上资源的供应与需求关系，包括网络流优化、中心定位与资源分配三个方面。其中定位问题是指已知需求源的分布，要确定最合适的供应点布

设位置；而分配问题是指已知供应点，要确定供应点的服务对象，或者说是确定需求源分别接受谁的服务。通常，这是两个需要同时解决的问题，合称为定位与分配问题。

假设研究区域内有 n 个需求点和 p 个供应点，每个需求点的权重（需求量）为 w_i，t_{ij}和d_{ij}分别为供应点j对需求点i提供的服务和两者之间的距离。如果供应点的服务能够覆盖到区域内的所有需求点，则：

$$\sum_{j=1}^{p} t_{ij} = w_i (i = 1 \sim n) \tag{21-11}$$

若规定每个需求点只分配给离其最近的一个供应点，则有：

$$\begin{cases} t_{ij} = w_i \cdots d_{ij} = \min(d_{ij}) \text{ 时} \\ t_{ij} = 0 \quad \text{其他情况} \end{cases} \tag{21-12}$$

网络整体的目标方程必满足：

$$\sum_{i=1}^{n} \sum_{j=1}^{p} c_{ij} = \min \tag{21-13}$$

其中，c_{ij}可以有以下几种基本理解（如图 21-4 所示）：

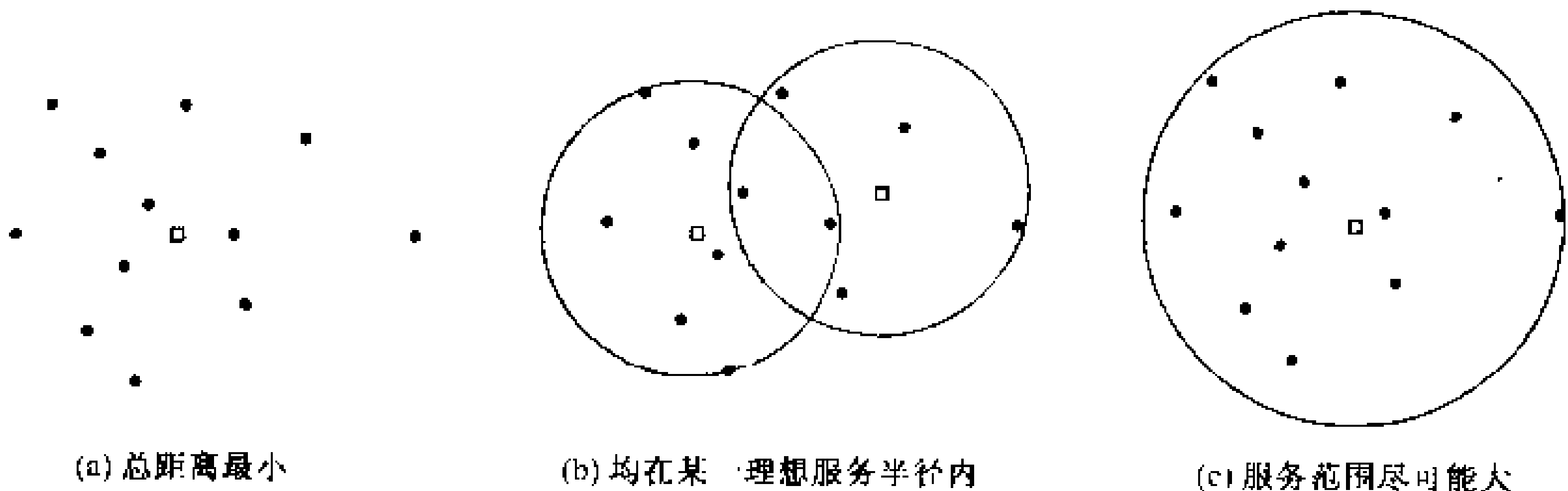


图 21-4　P-中心模型的基本形式

(1) 当要求所有需求点到供应点的距离最小时：

$$c_{ij} = w_i d_{ij} \tag{21-14}$$

(2) 当要求所有需求点均在某一理想服务半径 s 之内时：

$$c_{ij} = \begin{cases} w_i d_{ij} & d_{ij} \leqslant s \\ +\infty & d_{ij} > s \end{cases} \tag{21-15}$$

(3) 当要求所有供应点的服务范围尽可能最大，即新增需求点的代价最低时：

$$c_{ij} = \begin{cases} 0 & d_{ij} \leqslant s \\ w_i & d_{ij} > s \end{cases} \tag{21-16}$$

以上是资源分配问题的基本数学表达。在运筹学里，可以通过线性规划理论与方法求得其最佳解。但有一个问题，即当网络结点众多时（比如超过 100 个点），则计算量和需求量均非常大。因而，可以基于 GIS 的思想，根据地理网络的实际情况，用简洁图的方法来解决。

21.2.1 网络流优化算法

所谓网络流优化，是根据某种优化指标（如时间最少、费用最低、路径最短或运量最大等），找出网络物流的最优方案的过程。网络流优化分析的关键是要根据最优化标准扩充网络模型，即对结点、弧等地理要素进行性质细分和属性扩充。如结点可以细分为发货点和收货点；中心可以细分为发货中心和收货中心。

给定一个地理网络系统 D（V，E），c_{ij} 是表示边 e_{ij} 容量的一个非负数，c_i 是表示结点 v_i 容量的一个非负数，v_s 和 v_t 分别为其发点和收点。若 f 是关于网络边和网络结点的一个非负函数，即 $f_{ij}=f(e_{ij})$ 或 $f_i=f(v_i)$。如果函数 f 满足：

$$\begin{cases} f_{ij}=-f_{ji} \\ 0 \leqslant f_{ij} \leqslant c_{ij} \text{ 或 } 0 \leqslant f_i \leqslant c_i \end{cases} \tag{21-17}$$

则称 f 是网络系统 D 的一个可行流。由发货中心 v_s 到某一结点 v_t 的任意路径 P 的双向流之差称为函数 f 的一个流值：

$$v(f)_P = \sum_{e_{ij} \in P \in E} f_{ij} - \sum_{e_{ji} \in P \in E} f_{ji} \tag{21-18}$$

由发货中心 v_s 到收货中心 v_t 的任意路径的流值为：

$$v(f)_{st} = \sum_{(v_s, v_j) \in E} f_{sj} - \sum_{(v_j, v_s) \in E} f_{js} \tag{21-19}$$

其中，由发点 v_s 到收点 v_t 的所有路径中流值最大的即为该网络系统的最大流。实际地理网络中，寻找从某一点出发到另一点结束的最大流及其流向，对于交通运输方案的制定、物资紧急调运及管网路线的布设等十分有用。

定义：设 f 是地理网络系统 D（V，E）中的一个可行流。若 $f_{ij}=0$，则称弧 e_{ij} 为零流；否则为正弧。若 $f_{ij}=c_{ij}$，则称弧 e_{ij} 为饱和弧；否则为非饱和弧。若 P 为 D 中连接发货中心 v_s 与收货中心 v_t 的一条路径，若其前向弧为非饱和弧，后向弧为正弧，则称 P 为 D 中 f_{st} 的增广链。

求解最大流的增广算法的基本思想为：从 f_{st} 的任何一个可行流 f_1 开始寻找 f_{st} 的增广链：如果 D 中存在从 v_s 到 v_t 的 f_1 增广链，则对 f_1 进行增广，得到一个流值增大的可行流 f_2；然后再在 D 中寻找 f_2 的增广链 f_3……直到找不到流的增广链为止。此时的可行流就是所求的从 v_s 到 v_t 的最大流。

对上述算法略作修改就可以求解扩大网问题。具体做法是：对应新发点 S 到新收点 T，用无穷大容量弧（S，v_{s1}）、（S，v_{s2}），…，将新发点 S 与原发点 v_{s1}，v_{s2}，…，连接起来，并令新弧的距离为 1。增加了新发点、收点和新弧的扩大网具有以下特点：即新发点 S 到新收点 T 的任一个流对应于原有各发点至原有各收点的流；而且，新网的最大流对应于原网的最大流。因此，上述最大流算法可以应用于扩大网问题求解，并且由此产生的最大流必然生成原网络的最大流。

王家耀（2001）等学者相继对这一算法进行了研究、改进或技术实现。限于篇幅，此处不再赘述。

作为最大流问题的扩展，往往不仅要使网络的流量最大或达到预定值，还要求运送

成本最低，这就是最小费用流问题。设 w 是关于网络边和网路结点的一个费用函数，即 $w_{ij} = w\ (e_{ij})$ 或 $w_i = w\ (v_i)$，则地理网络系统 $D\ (V,\ E)$ 中 v_s 到 v_t 的一个可行流 f_{st} 的费用为：

$$w(f)_{st} = \sum_{e_{ij} \in E} w_{ij} f_{ij} + \sum_{v_i \in V \setminus \{v_s, v_t\}} w_i f_i \tag{21-20}$$

该网络系统 $D\ (V,\ E)$ 中 v_s 到 v_t 的所有可行流中费用最低的可行流即为所求的最小费用流。

21.2.2 定位与分配算法

许多资源分配问题的供应点布设需要满足多种组合条件，可以分解为多个单一目标进行求解。P-中心定位问题（P-median location problem）最初由 Hakimi（1964）提出，该模型假定结点代表需求点或是潜在地供应点，而弧段表示其间的通路或连接。Revelle 和 Swan（1974）将此问题表达为一个整数规划模型，其数学表达为：

假设研究区域内有 n 个需求点，现要从 m 个候选点中选出 p 个供应点为其服务，并要求供应点的服务能够覆盖到区域内的所有需求点，且使得服务总距离（或时间、费用等）最少。假设每个需求点的需求量为 w_i，t_{ij} 和 d_{ij} 分别为供应点 j 对需求点 i 提供的服务的系数（也称分配系数）和两者之间的距离，则定位要求为：

$$\sum_{i=1}^{n} \sum_{j=1}^{p} t_{ij} w_i d_{ij} = \min \tag{21-21}$$

$$\sum_{j=1}^{p} t_{ij} = 1 \cdots\cdots (i = 1 \sim n) \tag{21-22}$$

$$\sum_{j=1}^{p} \left(\prod_{i=1}^{n} t_{ij} \right) = p \cdots\cdots p \leqslant m \leqslant n \tag{21-23}$$

若规定每个需求点只由离其最近的一个供应点提供服务，则有：

$$\begin{cases} t_{ij} = 1 \cdots\cdots \text{当 } i \text{ 由 } j \text{ 服务时} \\ t_{ij} = 0 \cdots\cdots \text{其他情况} \end{cases} \tag{21-24}$$

若令 $E_{ij} = w_i d_{ij}$，则 E_{ij} 有以下几种基本理解：

（1）当要求所有需求点到相应供应点的距离均不超过 S 时：

$$E_{ij} = \begin{cases} w_i d_{ij} \cdots\cdots d_{ij} \leqslant S \\ \infty \cdots\cdots d_{ij} > S \end{cases} \tag{21-25}$$

（2）要求所选的供应点具有最大的服务范围，且需求点到相应供应点的距离均不超过 S 时：

$$E_{ij} = \begin{cases} 0 \cdots\cdots d_{ij} \leqslant S \\ w_i \cdots\cdots d_{ij} > S \end{cases} \tag{21-26}$$

（3）当要求供应点的服务范围最远不超过 T 时：

$$E_{ij} = \begin{cases} 0 \cdots\cdots d_{ij} \leqslant S \\ w_i \cdots\cdots S < d_{ij} \leqslant T \\ \infty \cdots\cdots d_{ij} > S \end{cases} \tag{21-27}$$

P-中心定位问题的解有以下 3 条性质：

(1) 每个供应点均位于其服务的需求点的中央；

(2) 所有的需求点均分配给离其最近的供应点；

(3) 若从最优解集中移去一个供应点，并用未选上的供应点代替，会导致目标函数的增加。

通常，有两种方法求解 P-中心问题，即最优化算法和启发式算法（Heuristic algorithms)。最优化算法目前还存在解决大型问题时能力不足；在众多的启发式算法中，交换式算法（Inter change）用得最多，其中包括著名的 Teitz-Bart 算法和 P. Densham 与 G. Rushton 利用点的空间邻接相关性的改进 Teitz-Bart 算法，即全局/局域性交换式算法（global and regional)。全局/局域性交换式算法充分利用了候选点和需求点的信息，计算效率大大提高。据 P. Densham 与 G. Rushton 的实验，该算法比 Teitz-Bart 算法效率提高 2～3.3 倍。该算法的实现过程为：

(1) 先任选 p 个候选点作为初始供应点集，将所有需求点分别分配到与之最近的供应点上，计算总的加权距离。

(2) 作全局性调整：① 从初始供应点集中选择一个供应点并删除它，使得将它删除之后的总加权距离增加得最少；② 从未选上的另外（m-p）个供应点中选出一个来代替被删除的供应点，使得总加权距离减少得最多；③ 如果②中总加权距离的减少值大于①中总加权距离的增加值，则用②中选择的供应点代替①中被删除的供应点，并更新总加权距离值，返回至①继续进行；否则，转入③。

(3) 作局域性调整：① 如果不是固定的供应点，则用其邻近的候选点来代替；② 若这一代替可以最大限度地减少总的加权距离值，则进行替换，直到 p-1 个供应点均被检验并无新的替换为止。

(4) 重复第（2)、第（3）步，直到这两步均无新的替换为止。

显然，经过全局调整和局部调整之后，求解的结果可以满足前述解的 3 条性质。尽管如此，启发式算法仍然存在一些不足（龚健雅，2001)：

- 不能保证解的结果一定最佳，但很接近；
- 不能平衡供应点之间的负担；
- 不限制供应点的容量；
- 初始供应点集的不同选择会对解的结果有影响。

§21.3 缓冲区分析算法

21.3.1 缓冲区的基本类型

缓冲区分析是 GIS 中使用非常频繁的一种空间分析技术，其实质是对一组或一类目标按某一缓冲距离（也称邻域半径）建立其周围缓冲区多边形图，然后将这一图层与目标图层叠加，进行分析而得到所需结果。显然，缓冲区分析不同于缓冲查询。前者涉及缓冲图层建立和叠置分析两个步骤，而后者只是检索缓冲区范围内所涉及的空间目标。叠置分析将在下节详细讨论，本节仅讨论缓冲区图层的建立及其有关算法问题。缓

冲区分析可分为点目标缓冲、线目标缓冲、面目标缓冲和复杂目标缓冲 4 种基本情况，其中复杂目标的缓冲区生成须先经过复杂计算和判断来生成一个复杂多边形或多边形集合。本节重点讨论点目标缓冲区、线目标缓冲区和面目标缓冲区的生成算法。

为讨论方便，首先定义以下几个概念：

(1) 轴线：即线目标的坐标点的有序串构成的迹线，或面目标的有向边界线。

(2) 轴线的左侧和右侧：沿轴线前进方向的左侧和右侧分别称为轴线的左侧和右侧。

(3) 轴线的凹凸性：轴线上顺序 3 点，用右手螺旋法则，若拇指朝上，则中间点左凹右凸；若拇指朝下，中间点左凸右凹；

(4) 多边形的方向：若多边形的边界为顺时针方向，则为正向多边形，否则为负向多边形。按公式 (18-14) 计算的正向多边形的面积为正，负向多边形的面积为负。

(5) 缓冲区的外侧和内侧：位于轴线前进方向左侧的缓冲区称为缓冲区的外侧，反之为内侧。

1. 点的缓冲区

点缓冲区的建立是以点状目标为圆心，以缓冲距离为半径所绘圆的区域。不同点状地目标的缓冲半径可以不一样。当两个或两个以上点状目标相距较近，或者其缓冲距离较大，则其缓冲区可能部分重叠。如图 21-5 所示。

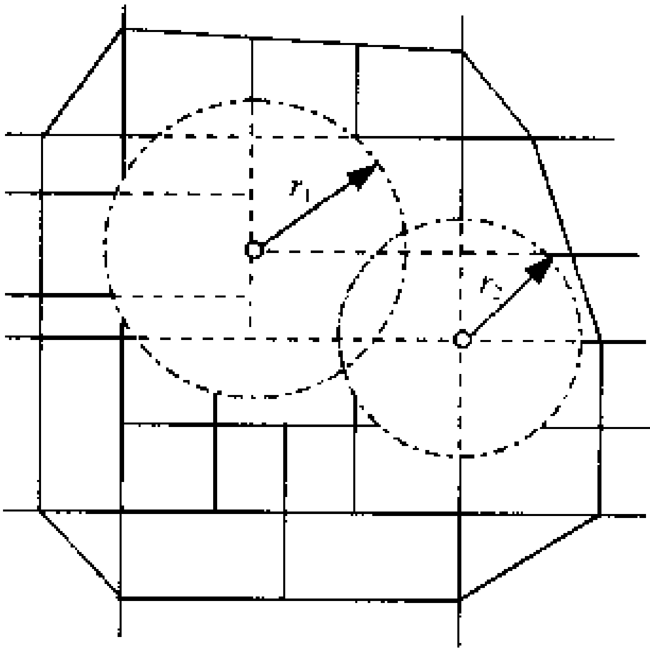


图 21-5　点的缓冲区部分重叠

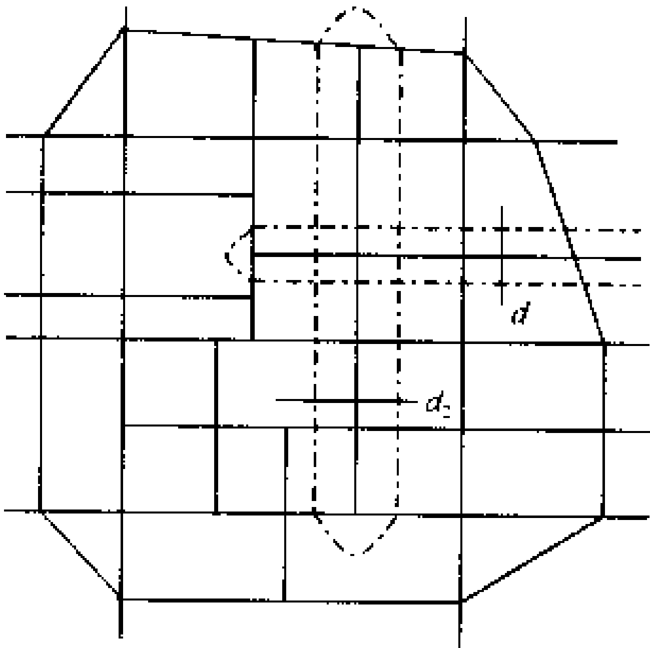


图 21-6　线的缓冲区部分重叠

点缓冲区的生成算法主要采用基于步进拟合的圆弧弥合法，即将圆心角等分为若干等份，用等长的弦来代替圆弧。已知半径为 R（缓冲距）的圆弧上一点 A (a_x, a_y)，按顺时针方向用 n 个等间距的离散点来逼近缓冲圆，即等间距圆心角 $\alpha = 360°/n$。设已知 OA 的方向角为 β，A 的顺时针方向的下一点为 B (b_x, b_y)，其坐标计算公式为：

$$\begin{cases} b_x = a_x\cos\alpha + a_y\sin\alpha \\ b_y = a_y\cos\alpha + a_x\sin\alpha \end{cases} \tag{21-28}$$

若按逆时针方向进行逼近，其坐标计算公式为：

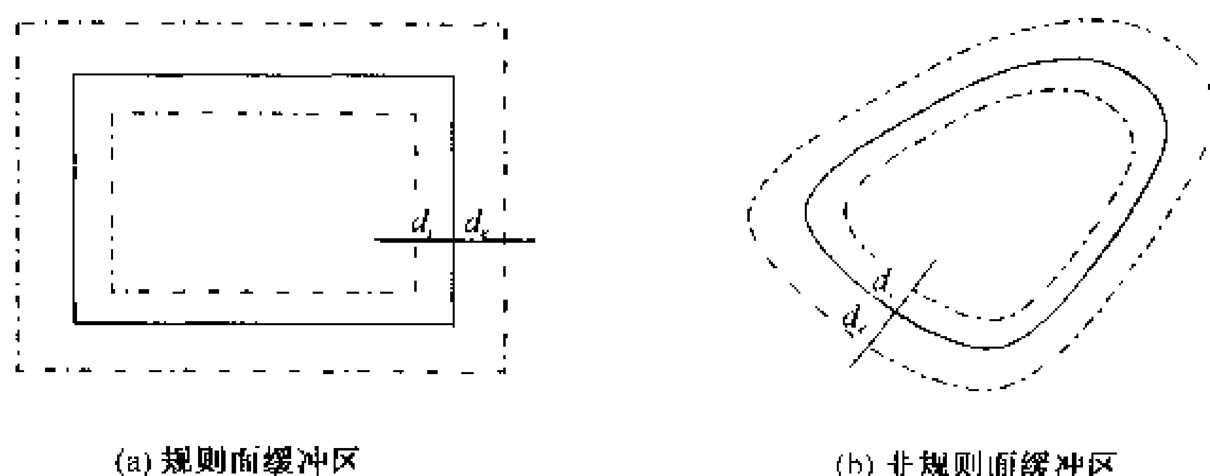
$$\begin{cases} b_x = a_x\cos\alpha - a_y\sin\alpha \\ b_y = a_y\cos\alpha + a_x\sin\alpha \end{cases} \tag{21-29}$$

2. 线的缓冲区

线缓冲区的建立是以线状目标为参考轴线，离开轴线向两侧沿法线方向平移一定距离，并在线端点处以光滑曲线（如半圆弧）连接，所得到的点组成的封闭区域即为线状目标的缓冲区。特殊情况下，可以指定不同线状目标的缓冲区宽度不一样，同一线状目标两侧的缓冲区宽度也可以不一样，甚至同一线状目标不同段的缓冲区宽度也可以不一样。同样，当两个或两个以上线状目标交叉、相距较近，或者其缓冲距离较大时，其缓冲区可能交叉或部分重叠。如图 21-6 所示。

3. 面的缓冲区

面目标缓冲区边界生成算法的基本思路与线目标缓冲区生成算法基本相同。所不同的是，面目标缓冲区生成算法是单线问题，即仅对非岛多边形的外侧形成缓冲区，对岛多边形的内侧形成缓冲区，而对于环状多边形的内外侧边界可以分别形成缓冲区。特殊情况下，可以指定不同的面状目标的缓冲区宽度不一样，甚至同一面状目标内外侧的缓冲区宽度也可以不一样，如图 21-7 所示。



(a) 规则面缓冲区　　(b) 非规则面缓冲区

图 21-7　面的缓冲区

21.3.2　缓冲线算法

由于线状目标和面状目标在 GIS 中是以坐标的形式存贮的。因此，生成线状目标和面状目标缓冲区的过程实质上是一个对线状目标上和面状目标边界线上的坐标点逐点求得其缓冲点的过程。其关键算法是缓冲区边界点的生成和多个缓冲区的合并。缓冲区边界点的生成有多种算法，代表性的有角平分线法和凸角圆弧法。

1. 角平分线法

以图 21-8 为例，基于角平分线的缓冲线生成的基本步骤如下：

(1) 选定线状或面状缓冲目标；

(2) 确定线状目标左右侧的缓冲距离（对于面状目标则为内外侧）d_l 和 d_r；

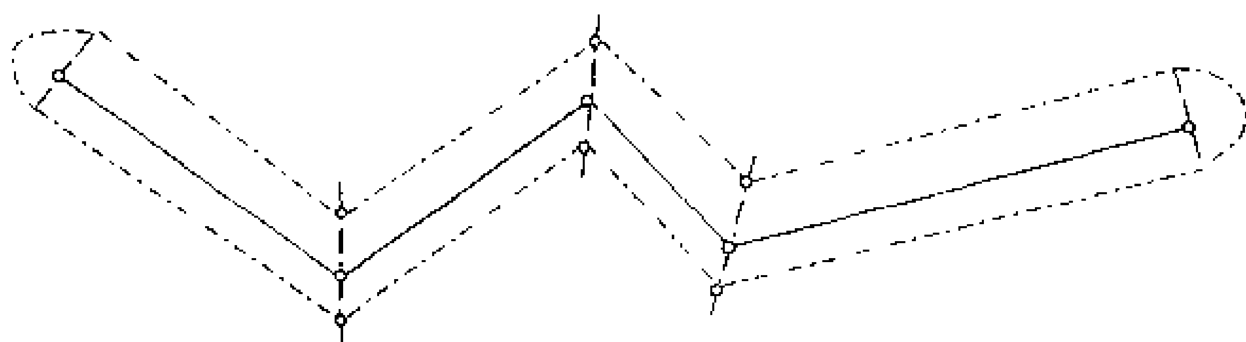


图 21-8　基于角平分线的缓冲线求解示例

(3) 提取线状目标的坐标序列；

(4) 沿线状目标轴线前进方向，依次计算轴线上各点的角平分线，线段起始点和终止点处的角平分线取为起始线段或终止线段的垂线；

(5) 在各点的角平分线的延长线上分别以左右侧缓冲距离 d_l 和 d_r，确定各点的左右缓冲点位置；

(6) 将左右缓冲点顺序相连，即构成该线状目标的左右缓冲边界的基本部分；

(7) 在线状目标的起始端点和终止端点处，以 (d_l+d_r) 为直径、以角平分线（即垂线）为直径所在位置分别向外作外接半圆；

(8) 将外接半圆分别与左右缓冲边界的基本部分相连，即形成该线状目标的缓冲区。

算法描述如下：

如图 21-9 所示，设轴线上顺序相邻的三点 A、B、C，其坐标依次为 (x_a, y_a)、(x_b, y_b) 和 (x_c, y_c)。令 AB、BC 连线的方位角为 α_{ab} 和 α_{bc}，沿前进方向左右侧的缓冲宽度分别为 d_l 和 d_r，则 B 点的左右缓冲点 B_l 和 B_r 的坐标分别为：

$$
\begin{aligned}
x_{bl} &= x_b - D_l \cos\beta_b \\
y_{bl} &= y_b - D_l \sin\beta_b
\end{aligned} \tag{21-30}
$$

$$
\begin{aligned}
x_{br} &= x_b + D_r \cos\beta_b \\
y_{br} &= y_b + D_r \sin\beta_b
\end{aligned} \tag{21-31}
$$

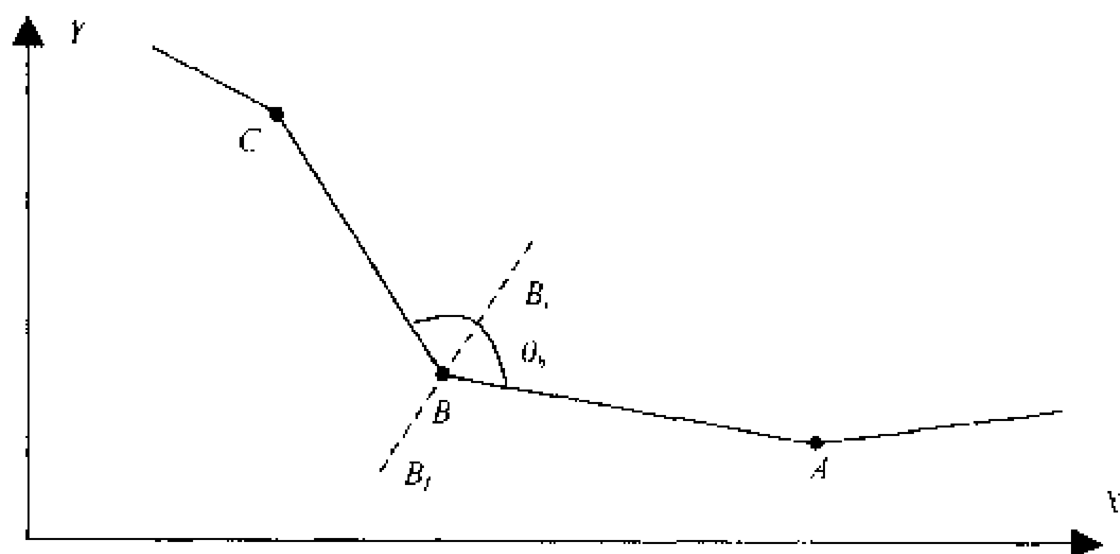


图 21-9　缓冲线求解示例

两式中：

$$D_l = \frac{1}{2\sin\left(\frac{\theta_b}{2}\right)} d_l$$

$$D_r = \frac{1}{2\sin\left(\frac{\theta_b}{2}\right)} d_r$$

$$\theta_b = \begin{cases} \alpha_{bc} - \alpha_{ba} \cdots\cdots\cdots\cdots\cdots\cdots (\alpha_{bc} > \alpha_{ba}) \\ \alpha_{bc} - \alpha_{ba} + 2\pi \cdots\cdots\cdots (\alpha_{bc} < \alpha_{ba}) \end{cases}$$

$$\alpha_{ba} = \begin{cases} \alpha_{ab} + \pi \cdots\cdots (\alpha_{ab} < \pi) \\ 2\pi - \alpha_{ab} \cdots\cdots (\alpha_{ab} \geqslant \pi) \end{cases}$$

$$\beta_b = \alpha_{ba} + \frac{1}{2}\theta_b$$

2. 凸角圆弧法

凸角圆弧算法（毋河海，1997）的基本思想是：在轴线的两端用半径为缓冲距的圆弧拟合；在轴线的各转折点，首先判断该点的凹凸性，在凸侧用半径为缓冲距的圆弧拟合，在凹侧用与该点关联的两缓冲线的交点为对应缓冲点。该算法的优点是可以保证凸侧的缓冲线与轴线等宽，而凹侧的对应缓冲点位于凹角的角平分线上，因而能最大限度地保证缓冲区边界与轴线的等宽关系。

1）轴线的凹凸性判断的矢量叉积算法

如图 21-10 轴线前进方向顺序三点 P_{i-1}（x_{i-1}，y_{i-1}）、P_i（x_i，y_i）和 P_{i+1}（x_{i+1}，y_{i+1}），矢量 $P_{i-1}P_i$ 与 P_iP_{i+1} 的叉积 S 为：

$$\begin{aligned} S &= P_{i-1}P_i \times P_iP_{i+1} \\ &= (x_i - x_{i-1})(y_{i+1} - y_i) - (x_{i+1} - x_i)(y_i - y_{i-1}) \end{aligned} \tag{21-32}$$

若 $S>0$，则 P_i 为凹点，三点为逆时针方向；

若 $S=0$，则该 3 点共线；

若 $S<0$，则 P_i 为凸点，该 3 点为顺时针方向。

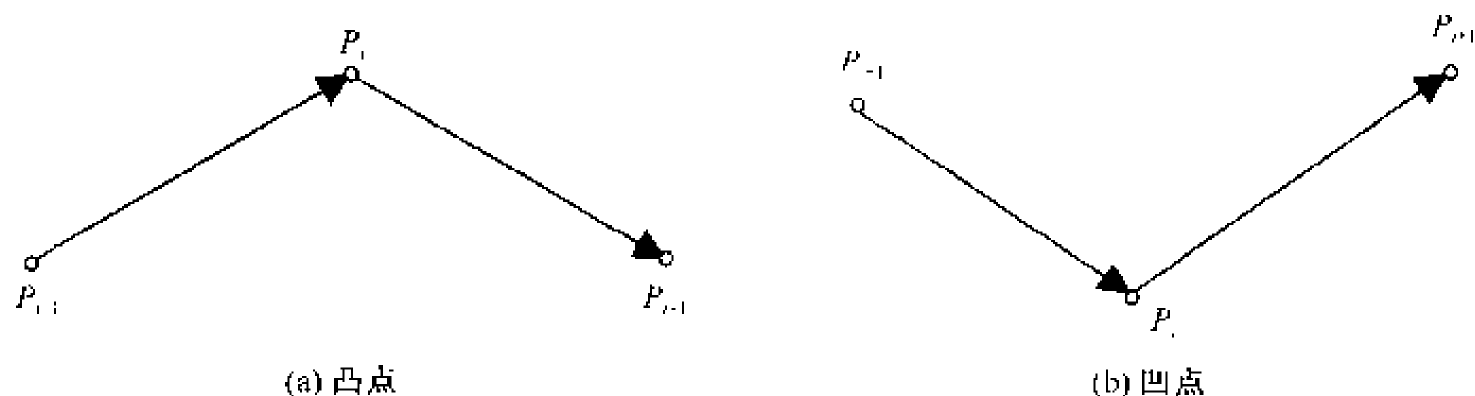


图 21-10　轴线的凹凸性判断

2）内侧缓冲点位的求解

如图 21-11 所示，将坐标原点平移至转折点 P_i，设此时线段 $P_{i-1}P_i$ 与 P_iP_{i+1} 的方

位角分别为 α_1 和 α_2，两左侧平行线交点 Q（x_q，y_q）到转折点的距离为 d，缓冲距为

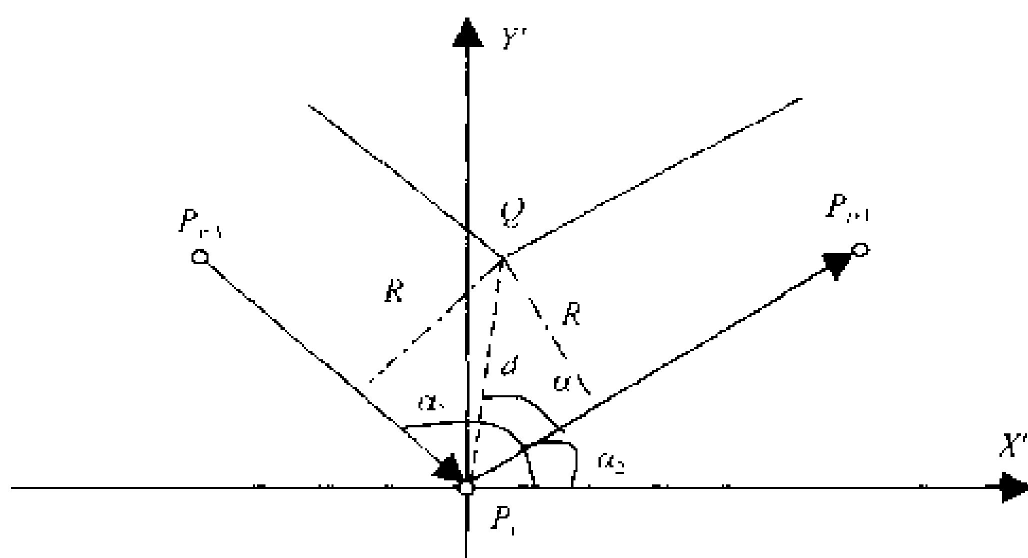


图 21-11　内侧缓冲点位的求解原理

R，令平分线的方向角为 α，则有

$$\begin{cases} R = d \times \sin(\alpha - \alpha_2) = y_q\cos\alpha_2 - x_q\sin\alpha_2 \\ R = d \times \sin(\alpha_1 - \alpha) = x_q\sin\alpha_2 - y_q\cos\alpha_1 \end{cases}$$

解之得：

$$\begin{cases} x_q = R(\cos\alpha_1 + \cos\alpha_2)/\sin(\alpha_2 - \alpha_1) \\ y_q = R(\sin\alpha_1 + \sin\alpha_2)/\sin(\alpha_2 - \alpha_1) \end{cases} \tag{21-33}$$

若坐标平移前 P_i 点的坐标为（x_i，y_i），则内侧交点的坐标为 Q（x'_q，y'_q）：

$$\begin{cases} x'_q = x_i + x_q \\ y'_q = y_i + y_q \end{cases} \tag{21-34}$$

同理，若缓冲线交点位于轴线右侧，则交点在坐标平移前的坐标为：

$$\begin{cases} x'_q = x_i - x_q \\ y'_q = y_i - y_q \end{cases} \tag{21-35}$$

3）外侧缓冲点位及方向的求解

如图 21-12 所示，圆弧弥合的起始点是转折点沿前一线段的法线方向向外平移一个缓冲距得到的点，终点则是转折点沿后一线段的法线方向向外平移一个缓冲距得到的点。在起始点与终止点相同的情况下，若圆弧弥合方向不同，将导致不同的结果。因此，有必要对圆弧弥合作方向限制，即沿轴线前进方向左侧弥合是顺时针方向，沿轴线

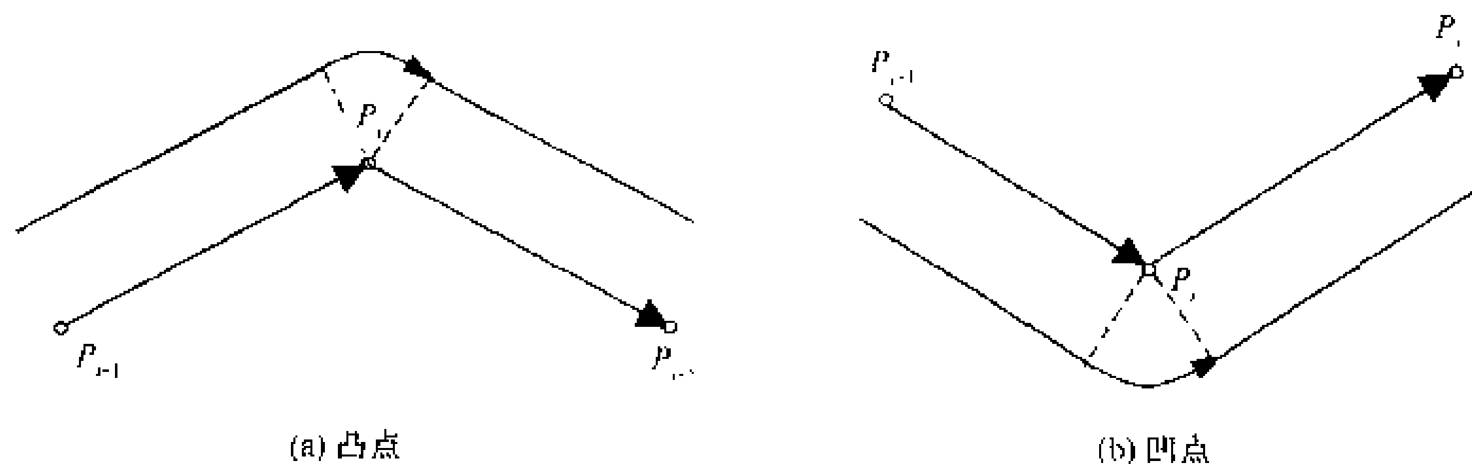


图 21-12　外侧圆弧弥合的位置与方向

前进方向右侧弥合是逆时针方向。

21.3.3 特殊情况处理

由于 GIS 中线状目标和面状目标的复杂性，在缓冲线生成过程中往往会遇到一些特殊情况，如缓冲线失真、缓冲线自相交和缓冲区重叠等

1. 缓冲线失真处理

当轴线转角太大时，会导致转角处的缓冲线交点随缓冲距的增大迅速远离轴线，出现尖角和凹陷等失真现象

如图 21-13 所示为按角平分线法得到的大转角处缓冲线。可见，由于 B 点的右转角太大，根据式（21 30）和（21 31）计算得到的 B_l 和 B_r 点均将远离 B 点，使缓冲区宽度发生变异，这是不合理的。处理措施为：

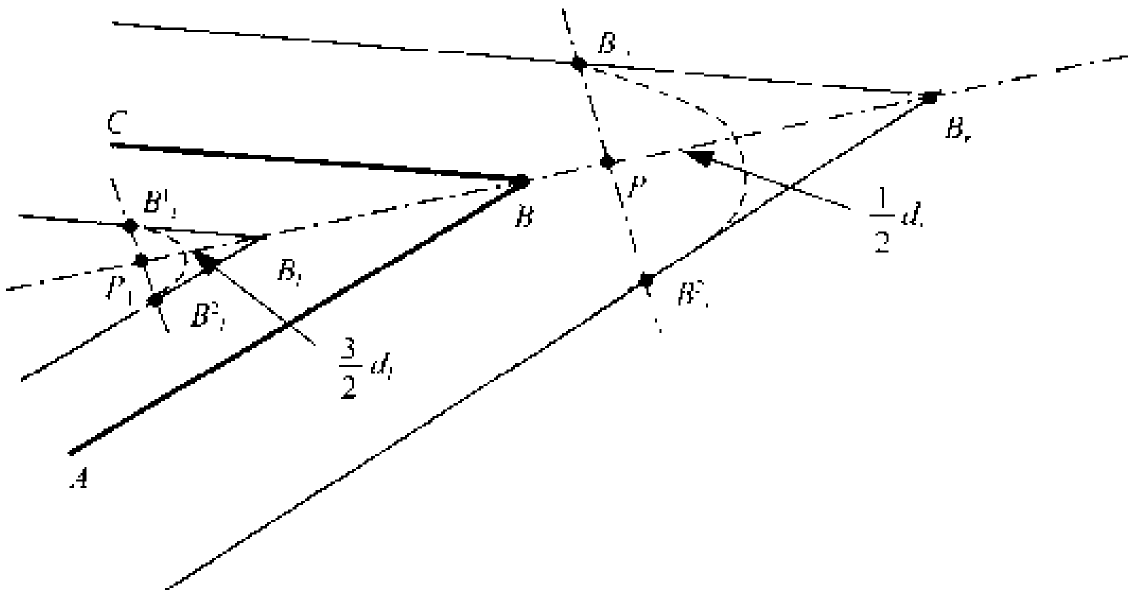


图 21-13 基于角平分线法的大转角处理

当 $D_r>0.75d_r$（或 $D_l>0.75d_l$）时，在 AB、BC 夹角的平分线上离 B 点 $\frac{1}{2}d_r$ 位置的 P_r 或 $\frac{3}{2}d_l$ 位置的 P_l 作其垂线，垂线与原缓冲线交于 B_r^1 和 B_r^2 两点和 B_l^1 和 B_l^2。分别以 P_r 和 P_l 为圆心、$B_r^1B_r^2$ 和 $B_l^1B_l^2$ 为直径，作圆弧分别连接 B_r^1、B_r^2 两点和 B_l^1、B_l^2 两点。此圆弧与原缓冲线相接即为修正后的缓冲区。

如图 21-14 所示为按凸角圆弧法得到的大转角处缓冲线。可见，由于 B 点的右转角太大，根据式（21-32）计算得到的缓冲线交点均将远离 B 点，与圆弧连接时出现尖角和凹陷，这也是不合理的。龚洁辉等（1998）通过大量实验和分析总结，将缓冲区边界失真现象分为 3 类 16 种，针对不同情况研究提出了不同的修正方法。

2. 缓冲线自相交处理

当轴线的弯曲空间不能容许缓冲区边界自身无压覆地通过时，缓冲线将产生自相交现象，并形成多个自相交多边形，包括岛屿多边形和重叠多边形。缓冲线自相交处理的关键是识别自相交产生的岛屿多边形和重叠多边形。

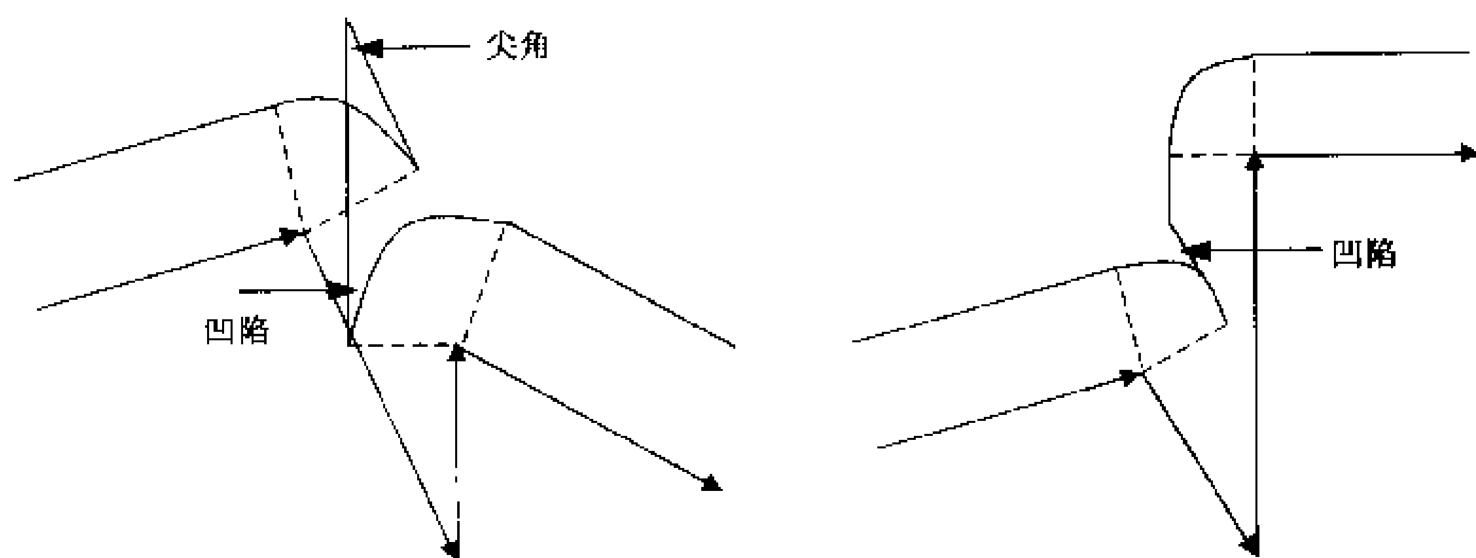


图 21-14　基于凸角圆弧的缓冲线失真

处理算法的基本思想为：求出缓冲线上的自相交点，判断这些点是入点（从缓冲区外侧进入内侧的点）还是出点（从缓冲区内侧出到外侧的点），并判断所产生的自相交多边形的性质。若是岛屿多边形则保留，否则保留面积最大的正向多边形为本缓冲区的外边界。

处理算法的基本步骤为：

1）判断自相交点是入点还是出点

据前述规定可知，所生成的缓冲线点串始终是顺时针方向，即轴线始终位于边界前进方向的右侧。同一相交点对于相交的两条缓冲线段具有相反的出入性质。可以利用判断轴线转折点凹凸性的矢量叉积方法来判断出入点的特性：求第一条线段的起点、交点 K 所确定的矢量与交点 K、第二条线段的终点所确定的矢量的叉积。若为正，则 K 是第 1 条线段的入点和第 2 条线段的出点；若为负，则 K 是第 1 条线段的出点和第 2 条线段的入点。如图 21-15 所示，可以判断出交点 K 分别是线段 12 的入点和线段 34 的出点。

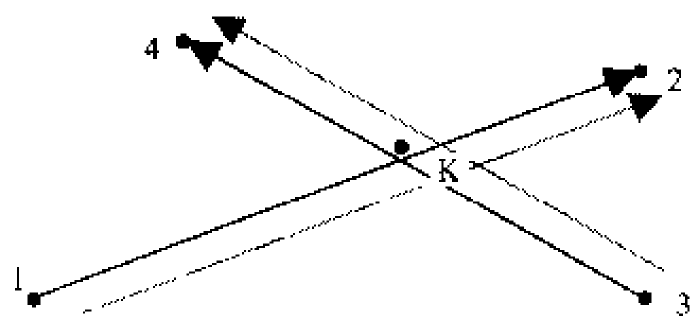


图 21-15　缓冲线交点的出入特性判断

2）判断自相交多边形是岛屿还是重叠区

由于所生成的缓冲线点串始终是顺时针方向的，故在边界自相交多边形中，负向多边形是岛屿；正向多边形中面积最大者是所求缓冲区的外边界，其他均为重叠区，要删除掉。如图 21-16 所示为一线状目标缓冲区处理前后的结果。

3．缓冲区重叠处理

缓冲区的重叠主要指不同目标的缓冲区之间的重叠。对这种重叠，首先要通过拓扑分析方法自动识别出落在某个特征区内部的线段或弧段，然后删除这些线段或弧段，则得到经处理后的相互连通的缓冲区，如图 21-17 所示。

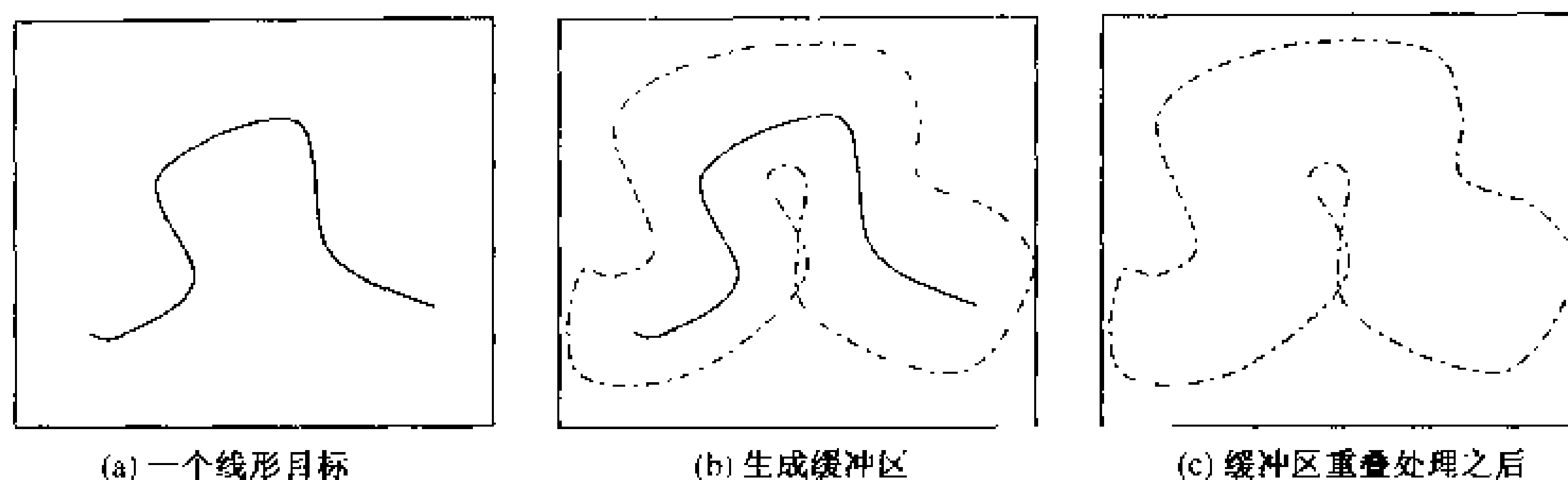


图 21-16　线状目标的缓冲线自相交处理

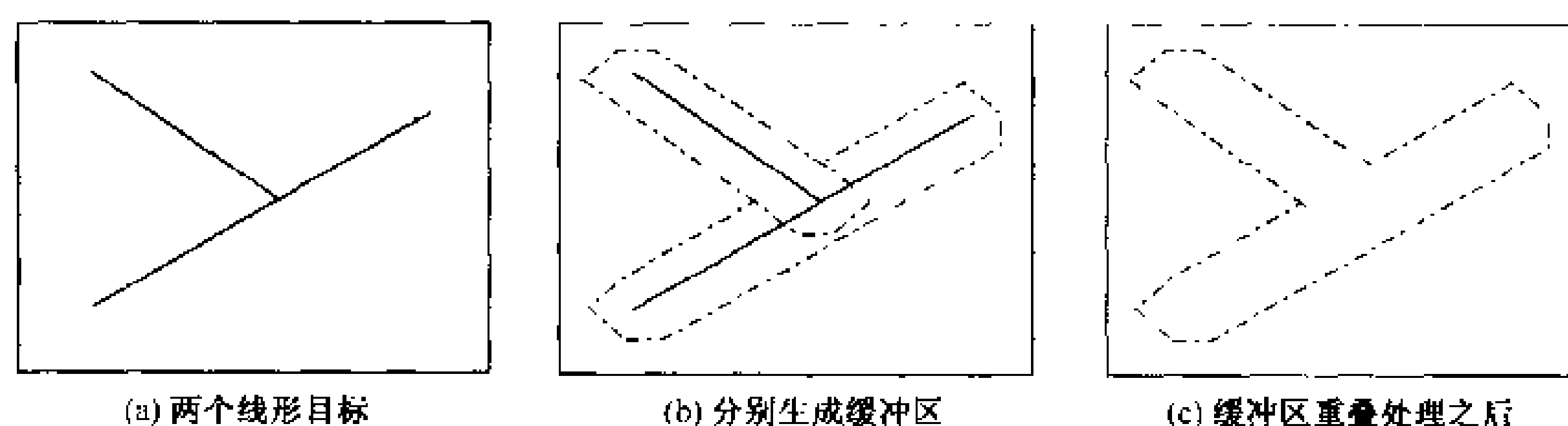


图 21-17　不同目标的缓冲区重叠处理

4. 动态缓冲区生成算法

动态缓冲区生成是针对两类特殊情况提出的：一类是流域问题，另一类是污染问题。在流域问题中，从流域上游的某一点出发沿流域下溯，河流的影响半径或流域辐射范围逐渐扩大；而从流域下游的某一点出发沿流域上溯，河流的影响半径或流域辐射范围逐渐缩小，类似问题还有参数动态变化的运动目标之影响范围分析。在污染问题中，污染源对邻近对象的影响程度随距离的增大而逐渐缩小，类似问题还有城市辐射影响分析、矿山开采影响分析等。

对于流域问题，可以基于线目标的缓冲区生成算法，采用分段处理的办法分别生成各流域分段的缓冲区，然后按某种规则将各分段缓冲区光滑连接；也可以基于点目标的缓冲区生成算法，采用逐点处理的办法分别生成沿线各点的缓冲圆，然后求出缓冲圆序列的两两外切线，所有外切线相连即形成流域问题的动态缓冲区。如图 21-18 所示。

针对污染问题，黄杏元等（1997）根据物体对周围空间影响度变化的性质，通过引入一个影响度参数，给出了 3 种动态缓冲区分析模型：

（1）物体对周围空间的影响度 F_i 随距离 d_i 呈线形衰减：

$$\begin{cases} F_i = f_0(1 - r_i) \\ r_i = d_i/d_0 \\ 0 \leqslant r_i \leqslant 1 \end{cases} \tag{21-36}$$

（2）物体对周围空间的影响度 F_i 随距离 d_i 呈二次函数衰减：

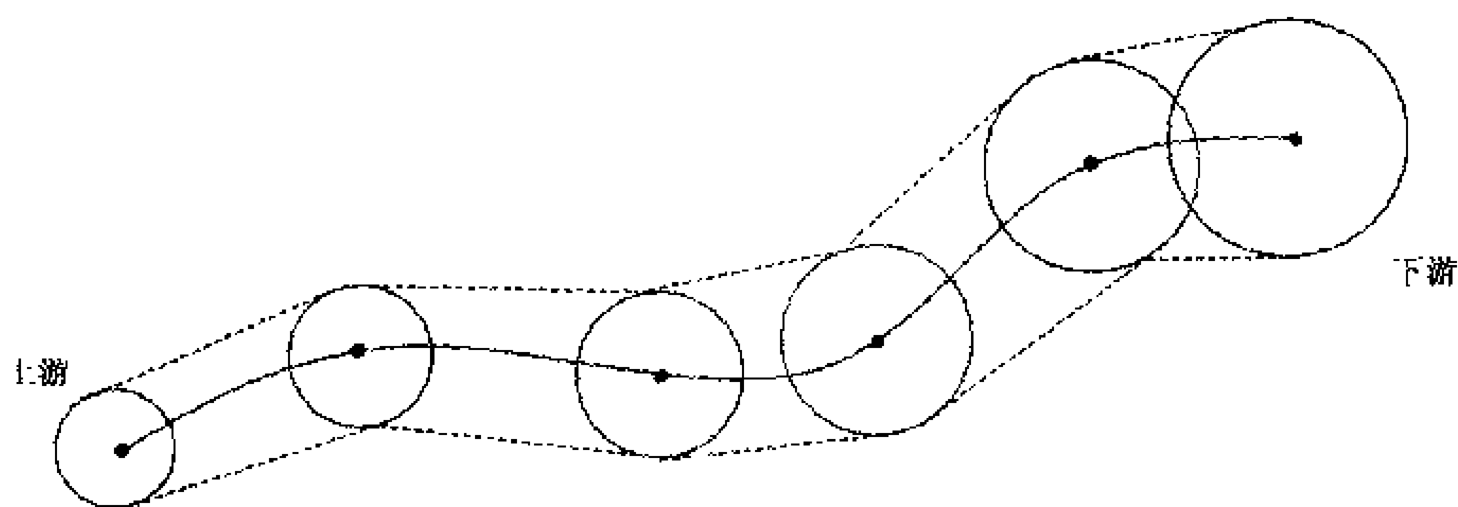


图 21-18　流域问题的逐点处理原理

$$\begin{cases} F_i = f_0(1 - r_i)^2 \\ r_i = d_i/d_0 \\ 0 \leqslant r_i \leqslant 1 \end{cases} \tag{21-37}$$

(3) 物体对周围空间的影响度 F_i 随距离 d_i 呈指数函数衰减：

$$\begin{cases} F_i = f_0^{(1-r_i)} \\ r_i = d_i/d_0 \\ 0 \leqslant r_i \leqslant 1 \end{cases} \tag{21-38}$$

以上三式中：f_0 为综合影响指数；d_0 为最大影响距离；d_i 为某点的实际距离；r_i 为距离比。

§21.4　叠置分析算法

叠置分析是 GIS 中的一项非常重要的空间分析功能，它是基于两个或两个以上的图层来进行空间逻辑的交、并、差运算，并对叠置范围内的属性进行分析评定。所涉及的图层中，至少有一个图层是多边形图层，称为基本图层，其他图层则可能是点、线或多边形图层。图层的数据结构可以是栅格的，也可以是矢量的，因而又可以将叠置分析划分为基于栅格的叠置分析和基于矢量的叠置分析两种类型。栅格叠置的结果是得到新的栅格属性；而矢量叠置实质上是拓扑叠置，其结果是新的空间特性和属性关系。

21.4.1　基于栅格的叠置分析算法

前已述及，栅格数据结构有非压缩和压缩两种存储形式，这将影响叠置分析算法

1. 非压缩的栅格图层叠置算法

对未经压缩的栅格图层进行叠置分析，其算法非常简单。设有两个未经压缩的栅格图层 A_{ij} ($i=1\sim m$；$j=1\sim n$)，B_{ij} ($i=1\sim m$；$j=1\sim n$)，要得到结果图层 C_{ij} ($i=1\sim m$；$j=1\sim n$)，只需对每个栅格元素进行逻辑运算，并将各像元的运算结果集合起来即可：

(1) 逻辑交

$$C_{ij} = A_{ij} \cap B_{ij} = \begin{cases} 1 \cdots A_{ij} = 1 \text{ 且 } B_{ij} = 1 \\ 0 \cdots A_{ij} = 0 \text{ 或 } B_{ij} = 0 \end{cases} \tag{21-39}$$

(2) 逻辑并

$$C_{ij} = A_{ij} \cup B_{ij} = \begin{cases} 1 \cdots A_{ij} = 1 \text{ 或 } B_{ij} = 1 \\ 0 \cdots A_{ij} = 0 \text{ 且 } B_{ij} = 0 \end{cases} \tag{21-40}$$

(3) 逻辑差

$$C_{ij} = A_{ij} - B_{ij} = \begin{cases} 1 \cdots A_{ij} = 1, B_{ij} = 0 \\ 0 \cdots A_{ij} = 0, \text{ 或 } A_{ij} = 1 \text{ 且 } B_{ij} = 1 \end{cases} \tag{21-41}$$

2. 压缩的栅格图层叠置算法

对采用行程、四叉树或二维行程编码的压缩图层，叠加分析时所采用的算法是不同的。此处讨论基于十进制 Morton 码（简称 M_D 码）压缩的栅格图层的叠置分析算法。

1) 子块逻辑运算基本规则

设有两个经 LQT 的 M_D 码压缩的栅格图层 A 和 B，要得到结果图层 C。以 M 表示图层某一子块（叶结点）的 M_D 码，M' 表示图层该子块的后继子块（叶结点）的 M_D 码减 1（即该子块右下角的地址码）。子块叠加运算的基本规则为：

规则 1（逻辑交）

$$M_A \cap M_B = \begin{cases} \begin{cases} M_C = 0 \\ M'_C = 0 \end{cases} & M'_A \leqslant M_B \text{ 或 } M'_B \leqslant M_A \\ \begin{cases} M_C = \max(M_A, M_B) \\ M'_C = \text{man}(M'_A, M'_B) \end{cases} & \text{否则} \end{cases} \tag{21-42}$$

规则 2（逻辑并）

$$M_A \cup M_B = \begin{cases} \begin{cases} M_C = \{M_A, M_B\} \\ M'_C = \{M'_A, M'_B\} \end{cases} & M'_A \leqslant M_B \text{ 或 } M'_B \leqslant M_A \\ \begin{cases} M_C = \min(M_A, M_B) \\ M'_C = \max(M'_A, M'_B) \end{cases} & \text{否则} \end{cases} \tag{21-43}$$

规则 3（逻辑差）

$$M_A \square | M_B = \begin{cases} \begin{cases} M_C = M_A \\ M'_C = M'_A \end{cases} & M'_A \leqslant M_B \text{ 或 } M'_B \leqslant M_A \\ \begin{cases} M_C = 0 \\ M'_C = 0 \end{cases} & M'_A \leqslant M_B \text{ 或 } M'_B \leqslant M_A \\ \begin{cases} M_{C1} = M_A, M'_{C1} = M_B - 1 \\ M_{C2} = M_B + 1, M'_{C2} = M'_A \end{cases} & \text{否则} \end{cases} \tag{21-44}$$

2) 压缩图层的叠置算法

数字图像或 GIS 中一个覆盖层中的任何一个图形（或区域）R 可以表示为一组 M_D

码（每个 M_D 码表示图形中的一个基本方块）的集合（两两互斥的并集）：

$$R = \{M_i\} = \sum_{i=1}^{n} M \qquad (21-45)$$

利用前述子块叠加运算的基本规则，定义压缩图层的叠加运算算法如下：

定理 1：两个图层 R_A、R_B 的交是其 M_D 码集合的交，并等于两两对应的 M_D 码交的集合，即：

$$R_A \cap R_B = \{M_{Ai} \cap M_{Bj}\} = \sum_{i=1,j=1}^{m,n} (M_{Ai} \cap M_{Bj}) \qquad (21-46)$$

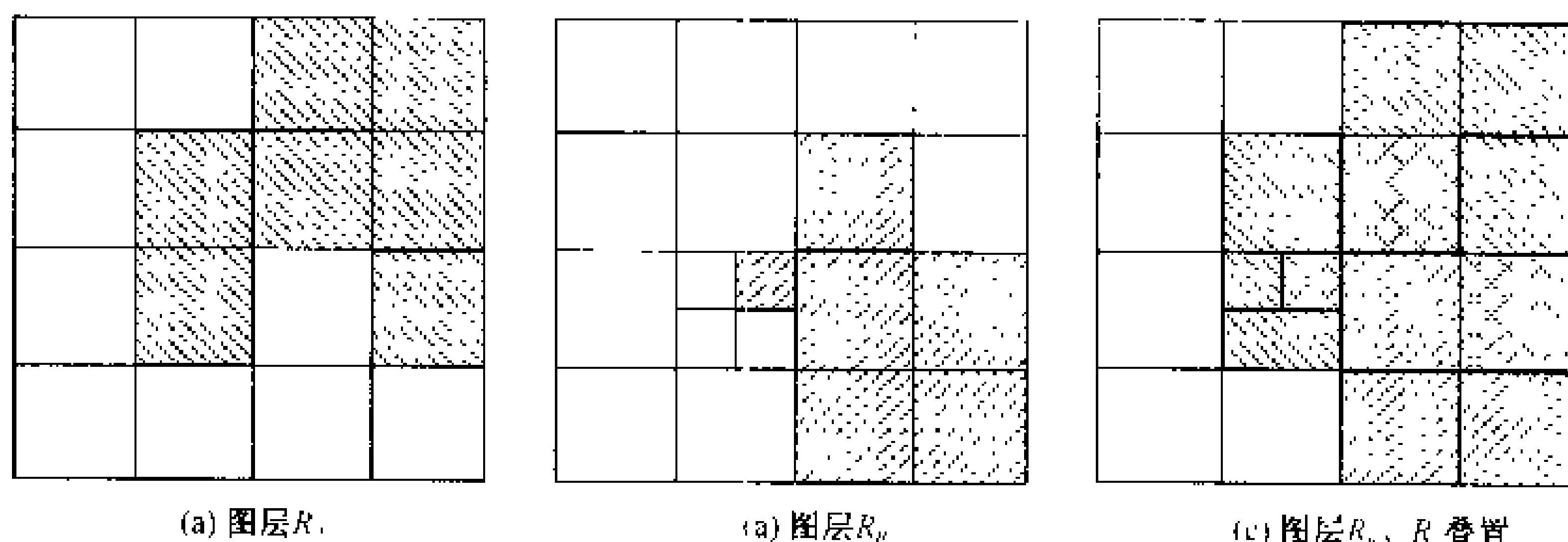
定理 2：两个图层 R_A、R_B 的并是其 M_D 码集合的并，并等于两两对应的 M_D 码并的集合，即：

$$R_A \cup R_B = \{M_{Ai} \cup M_{Bj}\} = \sum_{i=1,j=1}^{m,n} (M_{Ai} \cup M_{Bj}) \qquad (21-47)$$

定理 3：两个图层 R_A、R_B 的差是其 M_D 码集合的差，并等于两两对应的 M_D 码差的集合，即：

$$R_A - R_B = \{M_{Ai} - M_{Bj}\} = \sum_{i=1,j=1}^{m,n} (M_{Ai} - M_{Bj}) \qquad (21-48)$$

实际计算中，每一个基本方块或游程一般由两个 M_D 码确定，即该方块的起始 M_D 码和终止 M_D 码（即其后继子块的 M_D 码减 1）。以图 21-19 为例，两个图层 R_A 和 R_B 分别为：



(a) 图层 R_A　(a) 图层 R_B　(c) 图层 R_A、R_B 叠置

图 21-19　基于栅格的图层叠置分析示例

$R_A = \{(12, 15) \cup (16, 31) \cup (36, 39) \cup (52, 55)\}$

$R_B = \{(24, 27) \cup (37, 37) \cup (48, 63)\}$

根据定理 1 和规则 1、规则 2，可知：

$R_A \cap R_B = \{(12, 15) \cup (16, 31) \cup (36, 39) \cup (52, 55)\} \cap \{(24, 27) \cup (37, 37) \cup (48, 63)\}$

$= \{(12, 15) \cup (16, 31) \cup (36, 39) \cup (52, 55) \cap (24, 27) \cup (37, 37) \cup (48, 63)\}$

$= \{(0 \cup (24, 27) \cup 0 \cup 0) \cup (0 \cup 0 \cup (37, 37) \cup 0) \cup (0 \cup 0 \cup 0 \cup (52, 55)\}$

$$= (24, 27) \cup (37, 37) \cup (52, 55)$$
$$= \{(24, 27), (37, 37), (52, 55)\}$$
同理，根据定理 2 和规则 2，可知：

$R_A \cup R_B = \{(12, 15), (16, 31), (36, 39), (48, 63)\}$

同理，根据定理 3 规则 3，可知：

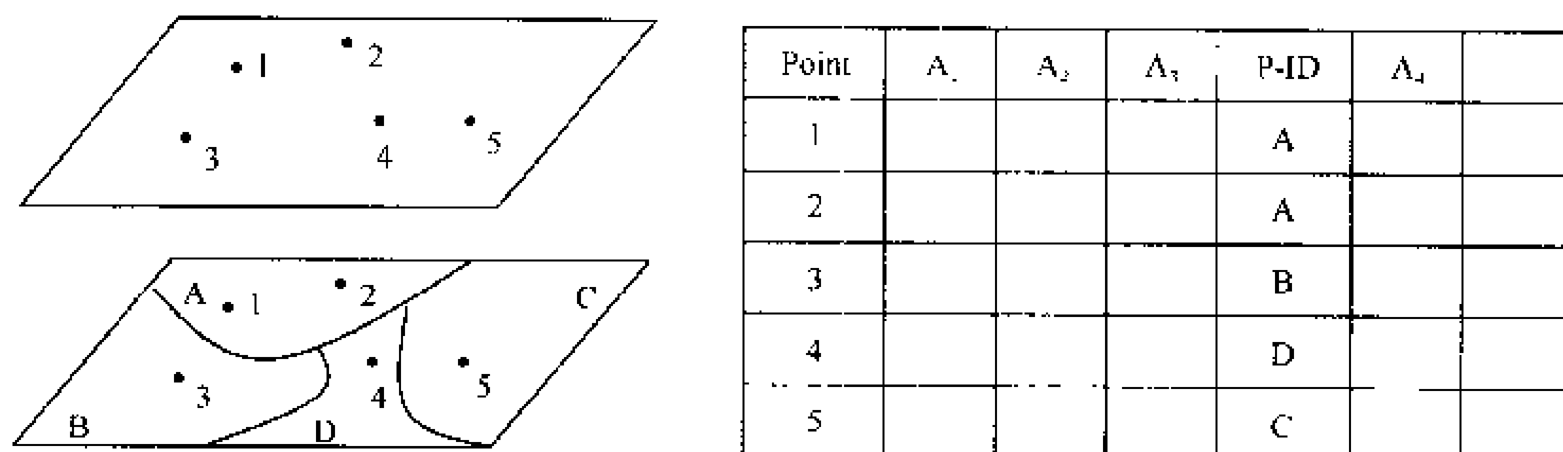
$R_A \quad R_B = \{(12, 15), (16, 23), (28, 31), (36, 36), (38, 39)\}$

21.4.2 基于矢量的叠置分析算法

随着计算机处理能力的提高和矢量叠置算法研究的成熟，基于矢量数据结构进行点与多边形、线与多边形或多边形与多边形叠置分析已成为可能，并在一些软件中实现（如 Geostar 等）。

1．点与多边形叠置（point overlay）

点与多边形叠置的实质是将一个含有点的图层叠置在一个多边形图层上，以确定每个点各落在哪个多边形内。该过程是通过点在多边形内的判别来完成的，通常得到关于点集的一个新的属性表，该表除包含点图层的原有属性之外，主要增加了各点所属多边形的目标标识。如有必要，还可以从相应的多边形属性表提取一些感兴趣的属性添加到点集的新属性表中。如图 21-20 所示。

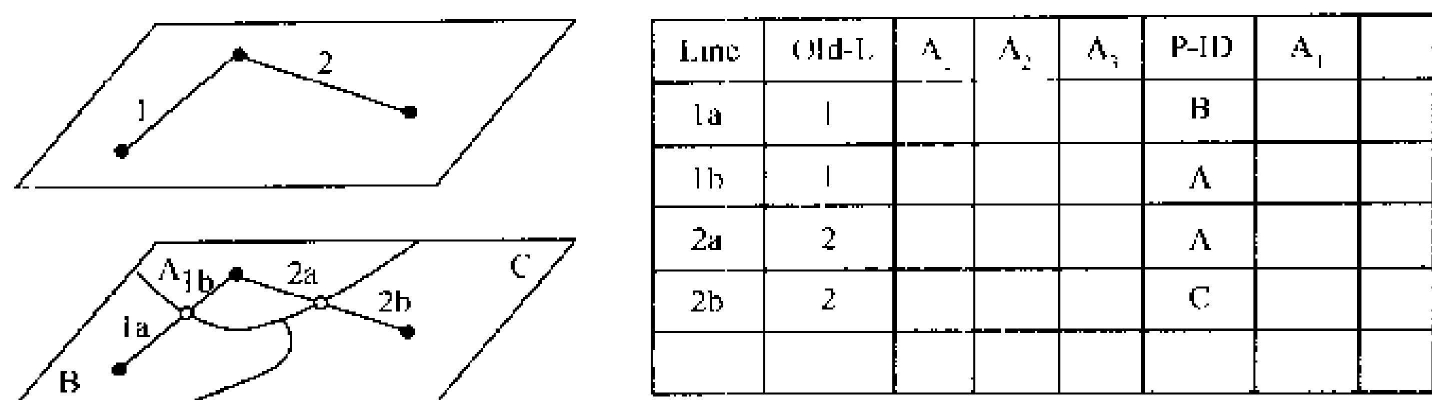


Point	A_1	A_2	A_3	P-ID	A_4	
1				A		
2				A		
3				B		
4				D		
5				C		

图 21-20 点与多边形叠置示例

2．线与多边形叠置（line overlay）

线与多边形叠置的实质是将一个含有线的图层叠置在一个多边形图层上，以确定每条线各落在哪个多边形内。该过程是通过线在多边形内的判别来完成。由于一个线目标往往跨越多个多边形，因而首先要进行线与多边形边界的求交，然后按交点将线目标进行分割，形成一个新的线状目标的结果集。基于该结果集可以得到线集的新的属性表，该属性表除包含线图层的原有属性之外，主要增加了分割后各线所属多边形的目标标识。如有必要，也可以从相应的多边形属性表提取一些感兴趣的属性添加到线集的新属性表中。如图 21-21 所示。



Line	Old-L	A	A_2	A_3	P-ID	A_1	
1a	1				B		
1b	1				A		
2a	2				A		
2b	2				C		

图 21-21　线与多边形叠置示例

3. 多边形与多边形叠置（Area overlay）

多边形与多边形叠置远比前两种复杂得多。这种叠置分析是指不同数据层或不同图层的多边形要素之间的叠置。就多边形的属性处理而言，可以分为合成叠置和统计叠置两种。合成叠置分析是指通过叠置形成新的多边形，并使新多边形具有多重属性，即需要进行不同多边形的属性合并，其方法包括取平均值、最大最小值或某种逻辑运算结果；而统计叠置分析是指确定一个多边形中含有其他多边形的属性类型的面积等参数值，即将其他多边形的属性提取到本多边形中来。两者的核心均是采用多边形与多边形裁剪算法形成新的多边形。

设两个原始多边形图层一个为本底多边形，另一个为上覆多边形，叠置得到的新多边形称为叠置多边形。首先要将两层多边形的边界全部进行求交运算和切割处理；然后根据切割的弧段重建拓扑关系，并产生一个叠置多边形图层，对其中的多边形重新编号；最后判断新叠置的多边形分别落在原多边形层的哪个多边形内，从而建立叠置多边形与原多边形的联系表。如果有必要，再从原多边形层抽取部分属性作为叠置多边形的属性。叠置过程如图 21-22 所示。

图 21-23 为两个多边形图层叠加的并、交、差结果。表 21-1 所示为空间叠置逻辑并的结果；表 21-2 所示为空间叠置逻辑交的结果，从表 21-1 中派生得到，即提取其中的交多边形 4、6 和 7；表 21-3 所示为空间叠置逻辑差的结果，也从表 21-1 中派生得到，即提出其中的差多边形 1、5 和 8。

表 21-1　多边形叠置逻辑并的结果

叠置 P	1	2	3	4	5	6	7	8	9	10
本底 P	A	0	0	A	A	A	A	A	0	0
上覆 P	0	b	a	a	0	b	c	0	b	c

表 21-2　多边形叠置逻辑交的结果

叠置 P	4	6	7
本底 P	A	A	A
上覆 P	a	b	c

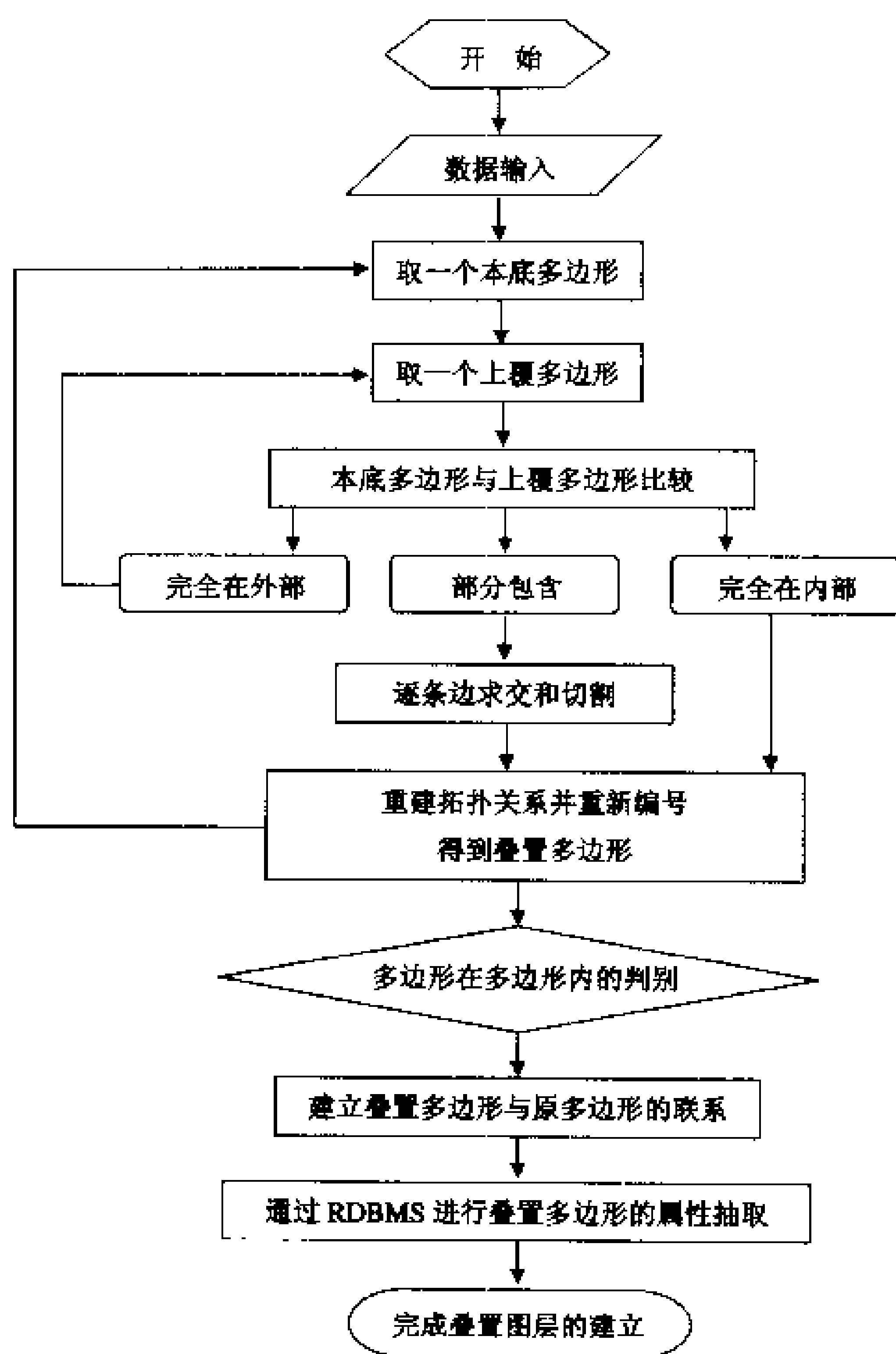


图 21-22 多边形叠置过程

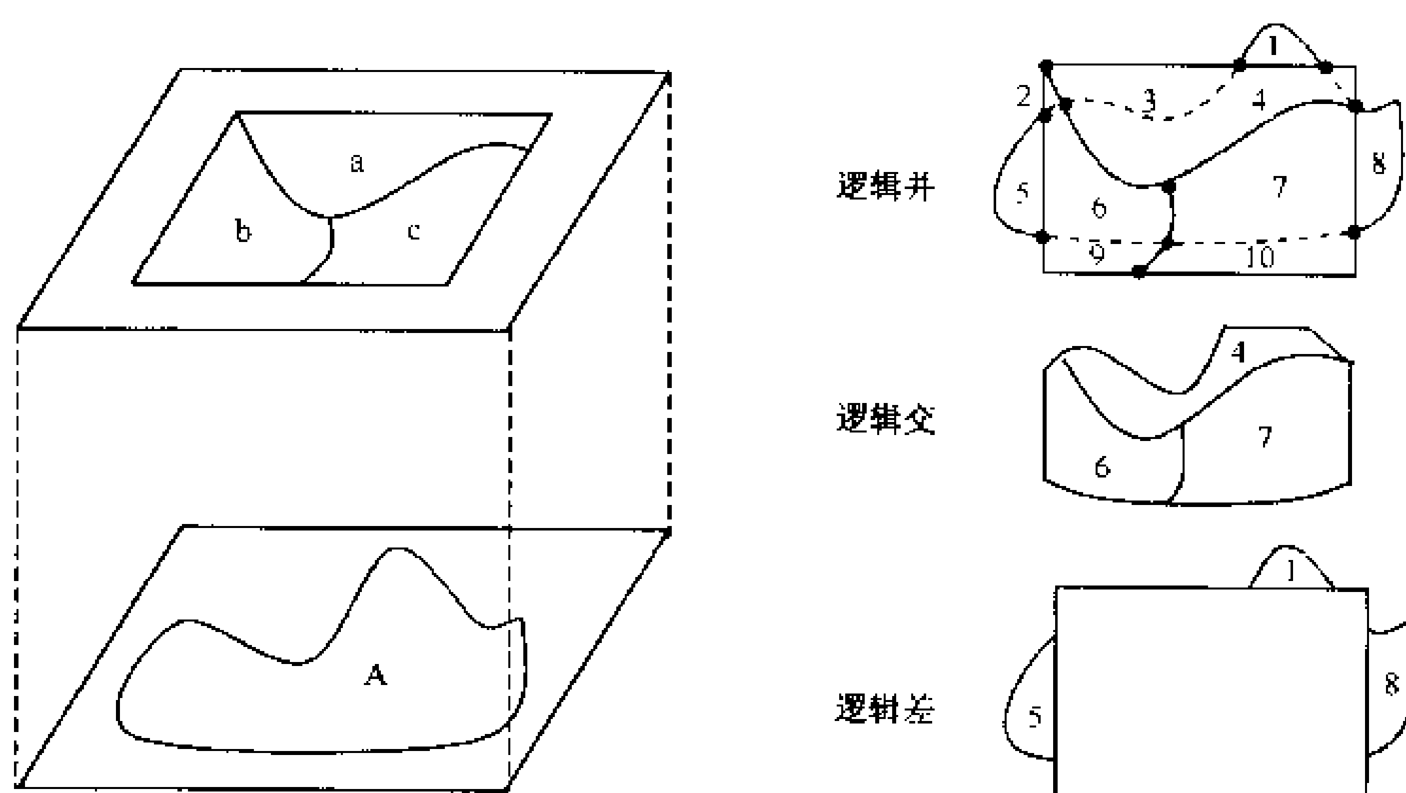


图 21-23 多边形叠置分析实例

表 21-3　多边形叠置逻辑差的结果

叠置 P	1	5	8
本底 P	A	A	A
上覆 P	0	0	0

参 考 文 献

丁跃民. 1999. GIS中实用空间算法设计的关键技术. 见龚健雅主编: 当代GIS的若干理论与技术. 武汉: 武汉测绘科技大学出版社. 114～121

龚健雅. 2001. 地理信息系统基础. 北京: 科学出版社. 232～253

龚治晖, 闫 玲. 1998. 确定地理网路中心服务的一种算法. 测绘学报. 27 (4): 357～382

郭仁忠. 1996. 空间分析. 武汉: 武汉测绘科技大学出版社

黄杏元, 汤 勤. 1989. 地理信息系统概论. 北京: 高等教育出版社. 122～138

黄杏元等. 1997. GIS动态缓冲带分析模型及其应用. 见: 中国GIS协会第三届年会论文集. 中国GIS协会主编. 116～121

沈永欢, 梁在中, 许履瑚, 蔡倩倩. 2002. 数学手册. 北京: 科学出版社.

王家耀. 2001. 空间信息系统原理. 北京: 科学出版社. 264～313

徐文华. 1989. 求解最短路径问题的一种计算机算法. 系统工程. 总第33期

徐业昌等. 1998. 基于地理信息系统的最短路径搜索算法. 中国图形图像学报. 3 (1): 39～42

Fischer, M., J. Henk, and R. Scholten and D. Unwin, eds., 1996. Spatial Analysis Perspectives on GIS. Taylor and Francis Ltd

第二十二章　GIS可视化操纵算法

可视化技术和虚拟现实技术的发展，推动了GIS由传统的2D地图的数字表达向3D可视化操纵的方向发展。“地理空间数据的可视化表达在一定意义上，可以被看成是数字时代的地图学”（Slocum T. A.，1999；高俊，2000）。可视化技术与GIS的集成是一个典型多学科交叉的前沿领域，其研究和发展需要多学科的共同推动。GIS可视化不仅是一个空间数据可视化的过程，还是一个艺术表达的过程，是数字环境下地球空间科学与计算机视觉艺术的结合。因而，GIS的可视化可以说是一个“如何创造性地描述客观世界的过程”（高俊，2000）。

近年来，3D可视化与交互操作（如3D空间查询与空间分析）已成为GIS的一个重要技术领域与应用方向，并逐步向虚拟现实和虚拟地理环境（林珲等，2001）方向发展，其中最突出的应用领域大到数字地球、数字流域，小至数字城市、数字矿山（吴立新等，2000，2001，2002，2003）和数字小区。为了提高3D漫游和交互操作速度，必须设计合理的数据结构与算法来管理和融合多源、多尺度、多精度（分辨率）的数据，包括DEM数据、数字正射影像（DOG）数据、线划矢量数据和各类属性数据等。当需要观察区域的整体全貌时，可以直接调用金字塔最底层的最低分辨率数据；若要了解局部细节，则可以根据区域大小和区域位置直接调用不同层次不同分块的数据。这一需求，对数据库管理技术、空间索引技术、并行计算技术、复杂模型简化技术、计算机性能、内存交换技术等均提出了更高的要求。

§22.1　地形简化算法

尽管许多学者已经围绕可视化技术与3DGIS的结合问题进行了探讨，但仍然没有形成一种通用的方法来连接GIS数据和可视渲染的内容（Gahengan M.，1999）。Gahengan M.（1999）指出，为了提高复杂场景的可视化速度和实现实时交互，需要解决以下问题：① 如何提高海量数据的3D漫游速度；② 如何提高3D模型可视化的真实感和美感；③ 用户的定位处理。在飞行或行走漫游的3DGIS或虚拟场景中，由于视点位置的不断变换，需要实时绘制不同场景的内容。从计算机硬件能力和内存开销考虑，通常可以采用降低可视场景的复杂度的方法，即减少每帧中需要绘制的多边形的数目，来提高海量数据的漫游速度。

为解决这一问题，细节层次模型（level of detail，LOD）具有普遍性和高效性，在飞行模拟和地形仿真中得到广泛应用。所谓LOD模型，是指根据不同的显示需求对同一个对象采用不同精度的集合描述。物体的细节程度越高，则描述得越精细，数据量越大；物体的细节程度越低，则描述得越简单，数据量越小。LOD模型实质上是按照一定的算法对原始几何模型进行简化，因此有时也称简化模型。它改变了传统的“图像质量越精细越好”的片面观点，参考人类视觉认知的模式，依据视线的主方向、视线在场

景中的停留时间、景物离视点的距离及景物在画面上投影区域的大小等因素，来综合决定景物应该选择的细节层次，以达到在保证实时图形显示的前提下，最大程度地提高视觉效果。

22.1.1 地形简化算法综述

地形简化其实包括地形细化（refinement）和地形简化（decimation）两个方面。地形细化是由粗到细的递归插入过程，即先用一组较少的点去拟合格网地形，然后根据精度和数据量的要求，再插入更多的点去拟合地形，如贪婪插入法（Greedy Insertion）；地形简化则是由细到粗的删减过程，即先由格网地形数据生成较详细的 TIN，在此基础上根据精度和数据量的要求再进行删减简化，如启发法（drop heuristic）。地形简化算法总结如表 22-1 所示。

表 22-1 地形简化算法汇总

大类	小类	提出者	技术特点
细化	并行贪婪法	Fowler R. J. et al (1979)	先用一个 2×2 的模板对格网地形进行滤波，选择保留点来形成初始 TIN；计算各三角形的最大误差点；将误差超限点（多个）插入到 TIN 中
	串行贪婪法	De Floriani et al. (1983)	扫描整个格网数据，找到一个与当前 TIN 绝对误差最大点，将该点插入到当前 TIN 中，时间复杂度为 $O(n^2)$
	相关贪婪法	Rippa S. (1992)	基于 Lawson（1977）的最大最小准则来确定相关三角形，进而决定点插入后两个三角形组成的四边形的对角线是否交换，故容易产生细长三角形
	改进的串行贪婪法	Garland M. and Heckbert P. S. (1995)	采用 Delaunay 三角剖分；时间复杂度为 $O[(m+n)\lg m]$
简化	启发式算法	Lee J. (1989)	先将格网地形的每个小矩形分成两个三角形，形成初始 TIN；然后利用节点的邻接点生成 Delaunay TIN；将误差大于阈值的结点删除，精度优于贪婪法
	等曲率算法	Scarlatos L. and Pavlidis T. (1992a)	通过收缩高曲率的三角形、合并共面相邻的三角形和交换边来实现 TIN 的简化
	边崩溃法	Hoppe H. (1998)	初始 TIN 的形成与启发式算法类似；以边的崩溃（edge collapse）进行简化，数据量大时，可将地形分块简化后合并
	重要点法	Chen Z. T. and Guevara J. A. (1987)	先用一个 3×3 的模板对格网地形进行滤波，将重要点保留下来生成 Delaunay TIN，非重要点删除，速度快，但无全局控制，对小范围的变换敏感

22.1.2 LOD 模型分类

Robin A. M. and Tom J. M. K.（1989）针对复杂场景与计算机实时处理能力的矛盾，为在计算机中以较少的时间及较高的图像质量绘制出复杂的场景，提出了复杂场景的层次表示方法及相关的绘制算法；潘志庚等（1996）从计算机图形学的角度，基于顶点的边界性与非边界性考虑和有序三角形环概念，以点到直线的距离和点到"平均平面"的距离为取舍准则，将 3D 三角化问题转化为 2D 带约束的三角化问题，进而提出了一种多层 TIN 的简化方法；孙敏等（2001）则将树木的分形模型、树木的图像模型与地形表面的纹理模型结合起来，对 3D 可视环境中的林木设计了 5 层次 LOD 模型。

纵观各种 LOD 模型，从几何结构上可以将其分为如图 22-1 所示的 3 种基本类型，即：① 不连续 LOD 模型；② 连续 LOD 模型；③ 结点 LOD 模型。

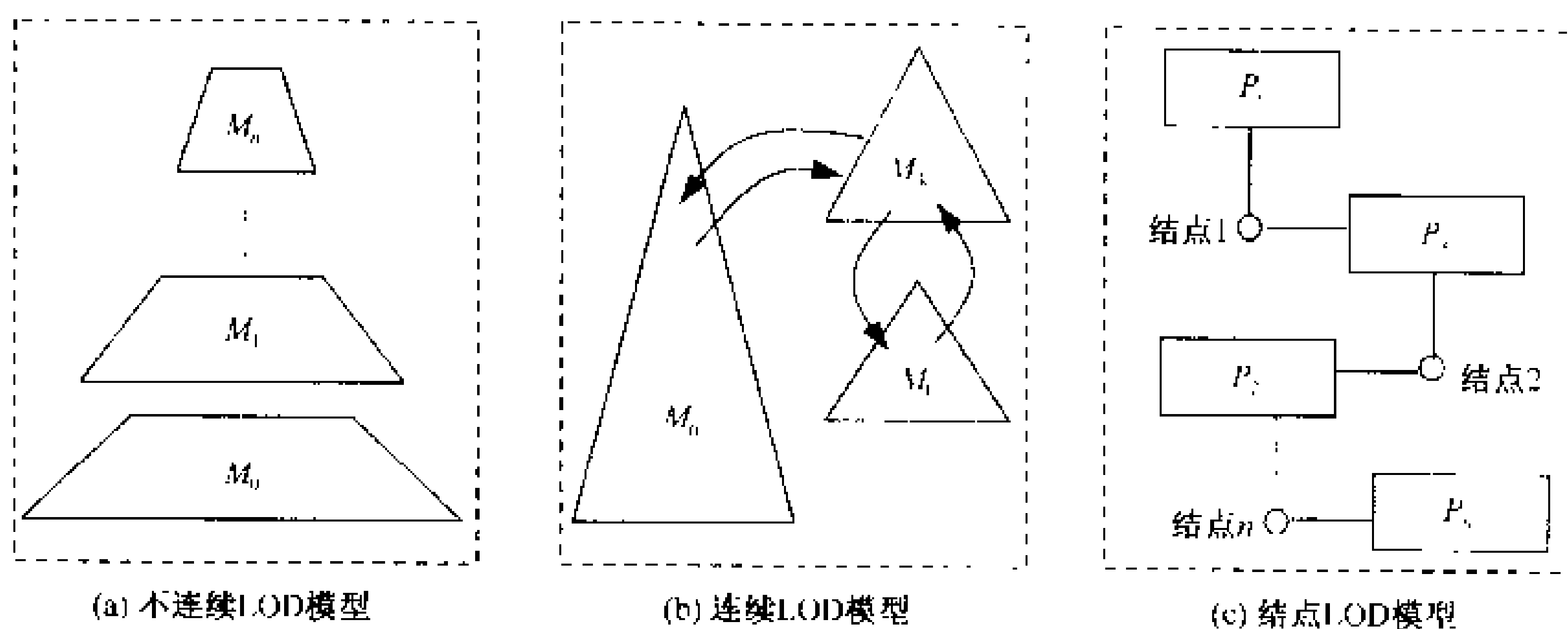


图 22-1 LOD 模型分类

（杨必胜，2002）

（1）不连续 LOD 模型：它实质上保存了原始模型的多个副本，每个副本对应某一特定的分辨率；所有副本构成一个金字塔模型。该模型的优点是不必在线生成模型，因而可视速度快；缺点是数据冗余大，容易引起几何数据的不一致性；而且由于不同分辨率之间没有任何关联，所以在可视化过程中，不同分辨率之间的转换容易引起视觉上的跳跃（popping）现象。

（2）连续 LOD 模型：它是在某一时间只保留某一分辨率的模型，在实际运用中根据需要，采用一定的算法实时生成另一分辨率的模型。该模型的优缺点正好与不连续 LOD 模型相反，即优点是没有数据冗余，能够保证几何数据的一致性和视觉连续性；缺点是需要在线生成不同分辨率的模型，算法设计复杂，可视速度慢。

（3）结点 LOD 模型：其本身是一个多分辨率结构，与不连续 LOD 模型不同的是，它的不同分辨率模型（部件）之间用节点相连，通过对节点的激活来操作相应的部件。当所有的节点均被激活时，它实质上就是一个全分辨率结构。该模型适合表达复杂的不连续的体模型对象，如建筑物类型。该模型结构在 VRML 语言、Multigen 系统和 Openflight 数据存储格式中得到广泛应用。

22.1.3 LOD 模型生成算法

围绕 LOD 模型的生成算法，国内外开展了广泛研究，尤其是关于大的连续曲面（如地形面）的 LOD 模型构建、复杂 CAD 的 LOD 模型构建和器官组织的 LOD 模型构建方面。以 CAD 模型和复杂曲面模型的 LOD 构模为例，所提出的算法有顶点杀死法（Schroeder W. et al, 1992）、面片移去法（Hamann, 1994）和面元合并法（Kalvin, 1996）等。所提出的各类 LOD 模型生成算法，根据其与视点、误差的关系和对拓扑的维护性能进行分类，如表 22-2 所示。

表 22-2　LOD 模型生成算法分类

LOD 模型生成算法	视点		误差		拓扑	
	相关	不相关	受控	不受控	保持	不保持

地形 LOD 模型的构造算法是较活跃的一个研究领域，在模型表达结构上一般采用层次结构、树结构和基于层次与树混合结构 3 种类型，如表 22-3 所示。

表 22-3　地形 LOD 模型生成算法分类

结构	算法	提出者	优点	缺点
层次结构	层次模型（hierarchical model）	De Floriani et al（1984）	能够控制模型的简化误差	不可避免产生狭长三角形 没有考虑视点的位置信息 有视觉跳动现象
层次结构	自适应层次模型	Scarlatos L. and Pavlidis T.（1992b）	部分避免产生狭长三角形	速度较低
层次结构	进度模型（progressive mesh, PM 法）	Hoppe H.（1996, 1998）	无视觉跳动现象 视相关	速度较慢
树结构	四叉树模型	Von H. B（1987）	层次清晰、结构规范 与空间索引统一 易构造与视点相关的模型	不适合非规则地形 可能出现区域边界不连续 可视化时可能出现空洞
树结构	二分树模型	Evans W. and Duchaineau M.（1997）	层次较清晰、结构较规范 易构造与视点相关的模型	不适合非规则地形 速度较慢 跳动现象依然存在
混合结构	层次模型-四叉树	杨必胜（2002）	能合理控制模型简化误差 适合不同区域的分别构建	需要解决分区边界裂缝的拼接问题

1. 层次结构的 LOD 模型自动生成算法

1）层次模型生成算法

De Floriani（1984）提出用层次结构描述地形 LOD 模型。在层次地形 LOD 模型中，

用一分三（ternary）的方法在 TIN 中移去或增加点，来构造层次 TIN 模型。设 τ_0 表示覆盖整个地形区域的根三角形；然后在 τ_0 中插入一个最大误差点 P，将 τ_0 分裂成 3 个新三角形；之后，判断所形成新三角形的误差是否大于该层次的误差阈值：如果是，则继续插入新点并对该新三角形进行分裂，直到所有新三角形的误差均小于该层次的误差阈值，即完成该层次的地形 LOD 模型的构建。此后，可以进行下一层次的 TIN 模型构建。该算法的时间复杂度为 $O(N_n)$（其中 N、n 分别为层数和节点数）。

2）自适应层次模型

Scarlatos L. et al.（1992）提出用自适应方法来建立层次 TIN 模型。与 De Floriani（1984）的三角形分裂规则不同，该算法将三角形的分裂分为 5 种情况。设三角形 t_i 的三个顶点为 V_1、V_2 和 V_3，所对的 3 条边分别为 e_1、e_2 和 e_3，令 t_i 和 e_1、e_2、e_3 的最大误差分别为 $E(t_i)$、$E(e_1)$、$E(e_2)$、$E(e_3)$，所对应的最大误差点分别为 p、p_1、p_2 和 p_3。若分裂误差阈值为 ε，则 5 种分裂情况如图 22-2 所示：

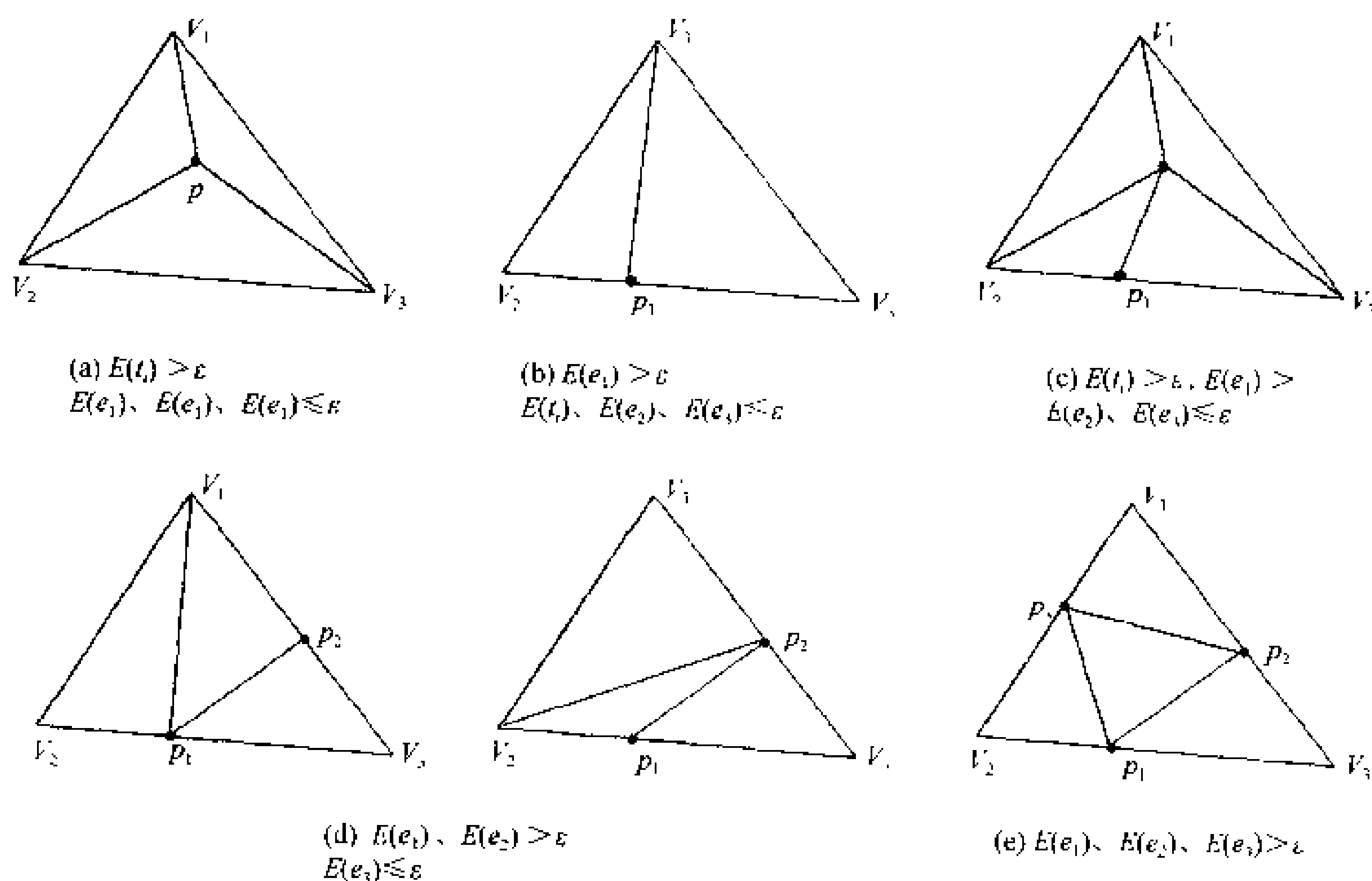


图 22-2　自适应层次模型三角形分裂规则

3）进度模型生成算法

Hoppe H.（1996）将进度格网（progressive mesh，PM）的概念引入地形 LOD 模型中，提出了进度模型生成算法（PM 法）。该算法通过一系列的顶点分裂（split）与合并（collapse）实现对原始模型的变换，从而构建不同分辨率的 LOD 模型，如图 22-3 所示。

TIN 顶点的邻接关系是 PM 算法的基础。如图 22-3 所示，顶点的分裂是通过增加一个顶点到原 TIN 中，从而产生更精细的 TIN 模型；顶点的合并则是通过删除一个顶点，从而产生更简略的 TIN 模型。即通过顶点的分裂与合并操作，得到一系列不同分

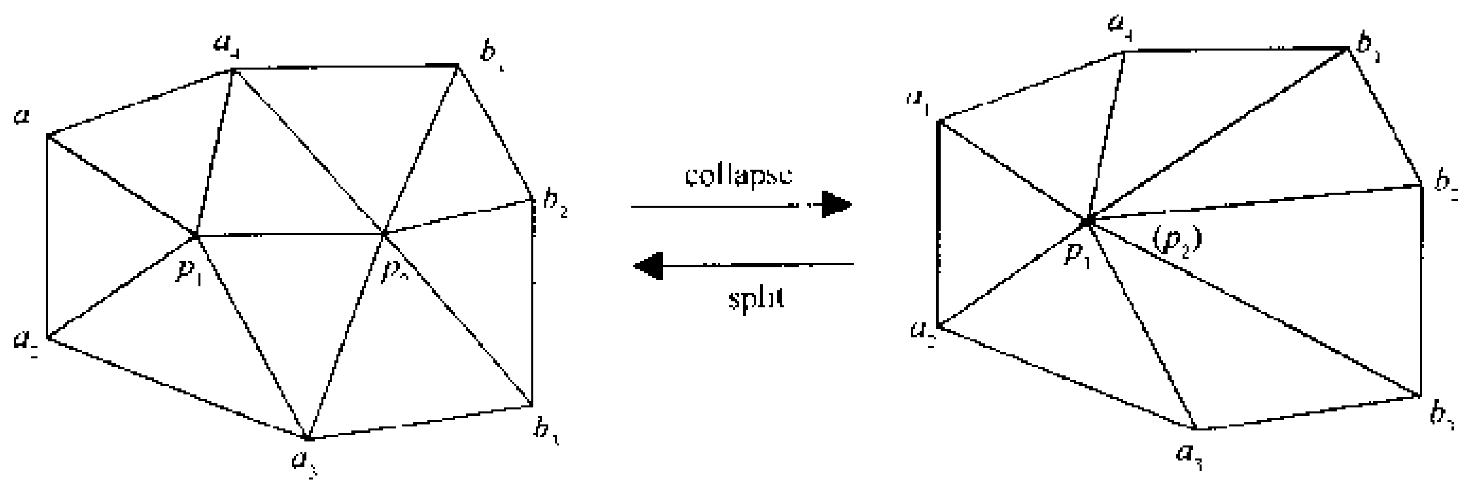


图 22-3　进度 LOD 模型

辨率的层次模型（LOD）。顶点的分裂与合并操作可以用下式描述：

$$M_0 \underset{\text{合并}}{\overset{\text{分裂}}{\rightleftharpoons}} M_1 \underset{\text{合并}}{\overset{\text{分裂}}{\rightleftharpoons}} \cdots \underset{\text{合并}}{\overset{\text{分裂}}{\rightleftharpoons}} (M_n = M) \tag{22-1}$$

以合并操作为例，一般分两步进行：第一步是确定崩溃边（edge collapse）的位置；第二步是保存边崩溃过程中顶点邻接关系的变化情况。崩溃边的选择和依据有多种，如 Hoppe H.（1996，1998）使用能量方程决定崩溃边；Xia J. C.（1996）使用顶点树决定崩溃边；Garland M. 和 Heckbert R. S.（1998）则使用边误差决定崩溃边。以使用边误差为例，Garland M. 和 Heckbert R. S.（1998）给出了边误差计算的详细公式，通过计算选择误差最小的边为崩溃边。在边的崩溃过程中，可以在崩溃边上及其邻近选择一个新的点来取代原崩溃边的两个顶点，也可以选择原崩溃边的两个顶点之一作为崩溃点。选择的依据是：如果两个顶点都是内部点，则选择误差较大的点作为崩溃点；如果其中有一个边界点，则选择边界点为崩溃点，如图 22-4 所示。在崩溃简化的过程中计算合并后的顶点误差：若小于给定的阈值则允许合并，否则不予合并。因此，该方法所构建的地形 LOD 模型具有全局统一的分辨率，模型质量较高，但是生成速度较慢。

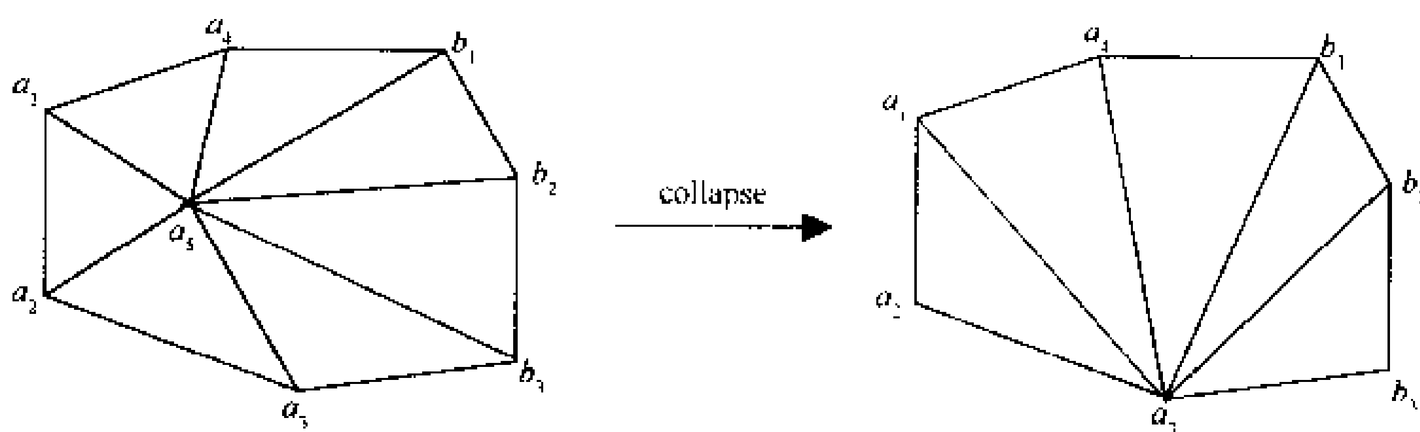


图 22-4　进度 LOD 模型的边界崩溃

为解决不同分辨率之间的跳动现象，Hoppe H.（1996）提出使用形态学（geomophics）变形的方法来实现不同分辨率地形 TIN 的切换和平滑过渡，即通过一个时间参数去控制三角形的变形。Hoppe H.（1998）还对 PM 法进行改进，提出了视相关 PM 法（vision dependence PM，VDPM 法）。VDPM 法的基本流程如下：

第一步：读入 TIN 模型，并建立顶点与三角形的邻接关系；

第二步：根据顶点与三角形的邻接关系，计算每个顶点和每条边的误差；

第三步：根据边误差建立边误差链表，并从小到大排队；

第四步：从边误差表中误差最小的边开始，依次进行边收缩，同时建立顶点层次树结构，并按顶点误差大小依次排列；

第五步：根据视点的位置依次判断顶点树中的顶点误差是否超出给定的阈值：若是，则对该顶点进行分裂操作；否则，进行顶点合并，同时更新顶点层次树结构。

第六步：重复第五步，直到满足要求为止。

2．树结构的 LOD 模型自动生成算法

1）四叉树模型生成算法

其实质是通过递归方法对地形进行自顶向下的四叉树分裂。在分裂过程中实时计算节点的误差：若该节点的误差大于规定的阈值，则该节点继续分裂，如此进行，直到所有节点的误差均小于给定的阈值。该算法要求地形格网的行列交点数满足 $2^n \times 2^n$，而实际上经常不能满足，则可能造成父节点的四个子节点所对应区域的分辨率不一样，从而在区域交界处出现不连续，可视化时出现空洞现象。

对此，可以采用限定四叉树（restricted quadtree）（Von H. B.，1987）的方法加以解决，即当某一区域与其相邻区域不连续时，通过对该区域的四叉树节点使用 6 种不同模式进行细分，以满足与邻接区域的连续性要求。由于该法基于四叉树对原始模型分块，因此与空间索引系统具有一定的统一性。

四叉树模型生成算法的基本流程为：

第一步：判断 Grid 模型的行列交点数是否满足$(2^n+1)\times(2^n+1)$的要求：若是，则进行下一步；若非，则对格网重新采样或补网；

第二步：对原始模型进行四叉树分裂，同时记录其 4 个子块的行、列数目，以及该顶点的误差；

第三步：进行递归运算，直到每个区域的长、宽的大小均等于原始 Grid 模型的最小分辨率时为止。

第四步：结束。

2）二分树模型生成算法

Duchaineau M. 等（1997）提出采用直角三角形的方法来建立层次网。该算法首先将正方形区域按对角线划分成两个直角三角形：若其误差大于给定的阈值，则使用直角顶点和斜边中点的连线将三角形分为两个三角形。重复以上过程，直到所有三角形的误差均小于给定的阈值。该法也可以推广至一般矩形区域，所不同的是使用大角（三角形中内角最大的角度，相等时任取其一）顶点和长边（三角形的边长最长者，相等时任取其一）中点的连线将三角形分为两个三角形。如图 22-5 所示。此时，需要在划分过程中判断三角形的大角和长边，因而速度较慢。

树结构算法的结构清晰，在 3D 可视化时，可以根据视点到地面的距离来确定不同区域需要绘制的层次。因而，该类方法容易构建出与视点相关的地形 LOD 模型，但仍

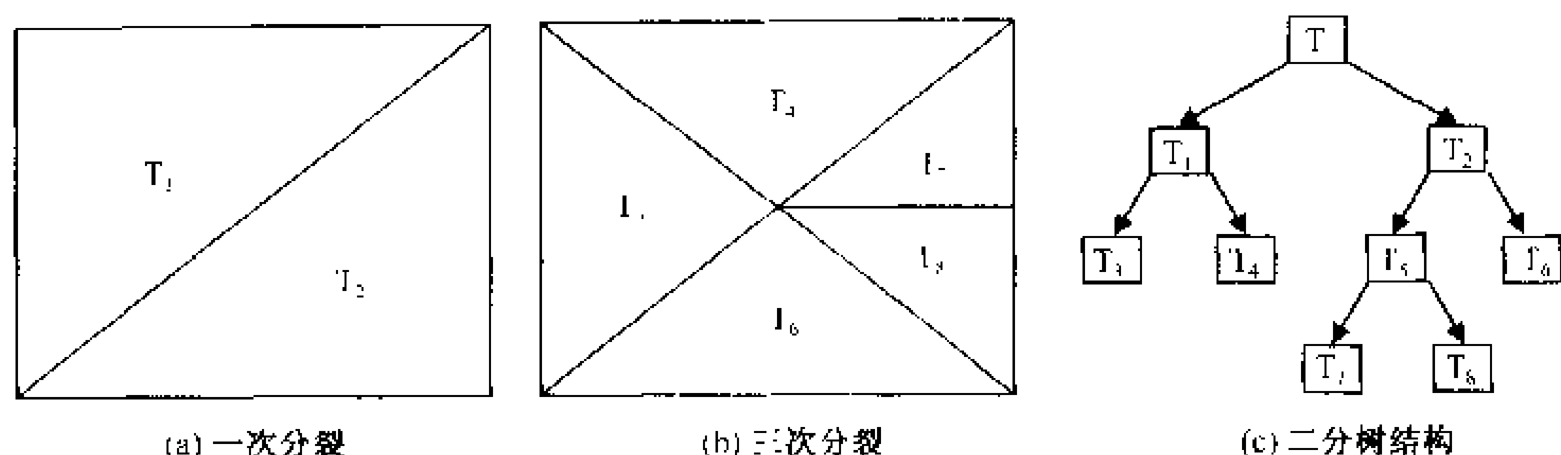


图 22.5　二分树 LOD 结构示例

然避免不了间跳现象。

在 GIS 可视化尤其是地形 3D 可视化中，无论采用层次结构模型还是树结构模型，首先要解决的是内存问题。据分析，这些算法消耗的内存均是原始格网 DEM 的 10～20 倍（王永明，2000）。尽管如此，由于地形的可视化过程受视点、视角、视线的限制，可视视野的范围往往是十分有限的，因此，与视点相关的算法是大型可视化的优先选择。此外，考虑到模型的质量，误差的受控性也需要关注，以确保可视化的真实性。

3．混合结构的 LOD 模型生成算法

杨必胜（2002）基于“分治”思想和混合结构，将原始建模区域 M 分成 n 个子块（M_1，M_2，…，M_n，$M=\bigcup_{i=1}^{n}M_i$），各子块 M_i 分别采用 TIN 模型或 Grid 模型来独立构建 LOD 模型 M_i'，然后对构造结果进行合并，生成最后的全区域 LOD 模型 $M'=\bigcup_{i=1}^{n}M_i'$。基于混合数据结构描述的地形 LOD 模型在 XOY 平面坐标系中建立了一定大小的格网索引，且每个格网的索引记录了该格网的属性是 TIN 还是 Grid（0 表示 Grid，1 表示 TIN）。

杨必胜（2002）设计的算法实现步骤如下：

第一步：将建模区域 M 划分为 n 个子块；

第二步：依次判断各子模块 M_i 的属性特征，若是 Grid 则执行第三步；若是 TIN 则执行第四步；

第三步：根据 Grid LOD 模型的生成算法构建 Grid LOD 模型，计算并判断节点的误差：若误差在给定的阈值之内，则停止构建；否则继续构建，直到满足条件得到 M_i' 为止，并转向第五步；

第四步：根据 TIN LOD 模型的生成算法构建 TIN LOD 模型，计算并判断节点的误差：若误差在给定的阈值之内，则继续简化；否则简化结束，得到最后的 M_i'；

第五步：重复第二步，直到所有的子块的 LOD 模型均构造完毕；

第六步：合并所有子块的 LOD 模型，得到区域 M 的最终 LOD 模型 $M'=\bigcup_{i=1}^{n}M_i'$；

第七步：算法结束。

该算法通过对原始大区域进行分块，减少了直接进行 LOD 建模的区域数据量，一定程度上可以提高算法的运行速度，尤其为并行计算提供了基础。但是，也正因为区域

分块，由于不同区域离视点的远近不同而分辨率不同，导致在3D可视化时在不同模型（Grid和TIN）的子块的接边处出现“裂缝”，必须进行处理，如图22-6所示。

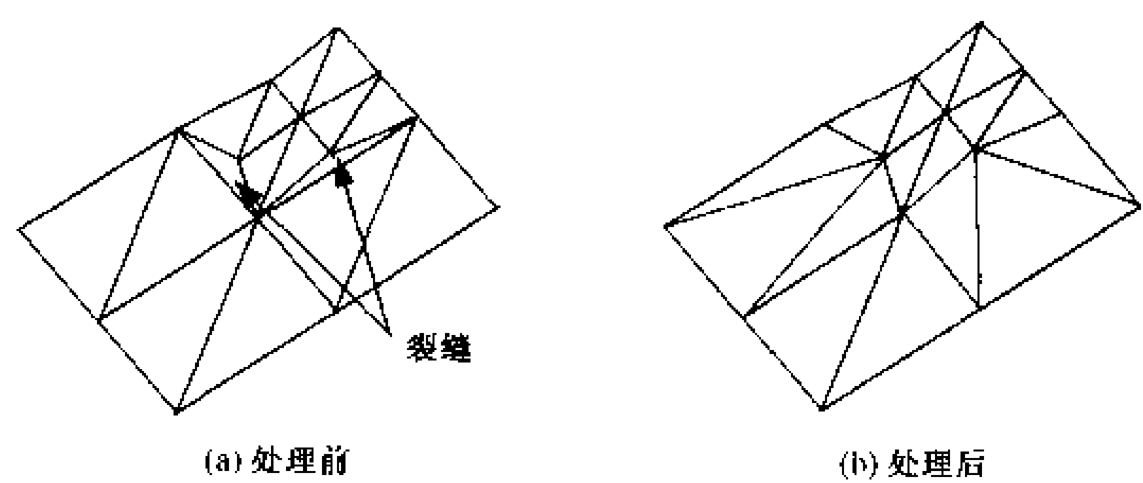


图22-6 混合模型子块接边处的裂缝与处理结果

§22.2 多分辨率纹理生成算法

以影像数据作为纹理特征来增强图形的真实感和提高丰富度，在3D可视化过程中对于弥补3D模型的几何数据的描述不足和提高可视化效果具有重要意义，并且可以在很大程度上减轻图形硬件的负担，提高图形渲染的速度。但同时，由于影像数据占用较大的内存空间（以RGB为例，每个像元占用3个字节），一般的计算机图形渲染设备限制了单次装载影像的大小，如目前OPENGL渲染机支持的单张影像的最大范围为2048×2048。而实际情况下，地形与城市建筑影像的范围远远大于这一规模，必须采用特殊的处理技术来解决这一矛盾，即根据纹理分辨率的视点相关性来生成多分辨率的纹理。

22.2.1 一般算法

为解决纹理影像与计算机的渲染机处理能力之间的矛盾，人们研究提出了多种解决的技术方法。其中，最具代表性的是生成多分辨率纹理的Mip-map技术（又称金字塔技术）和Clip-map技术。

1. 金字塔（Mip-Map）算法

当景物多边形由近及远或由远到近进行变换时，其投影到屏幕上的面积也会发生相应变化，导致一个屏幕像素可能对应多个多边形（距离较远时）或多个屏幕像素对应一个多边形（距离较近时）。若多边形是贴了纹理的，则相应会引起该屏幕像素对应多个纹理元素（texel），或反之。如此，必然造成纹理闪烁或马赛克等失真效果。目前的图形渲染设备普遍采用Mip-map技术来处理一个屏幕像素对应多个纹理元素的情况。

该算法的基本思想是采用不断降低纹理分辨率的办法将原纹理的数据规模缩小，使得最终1个屏幕像素仅对应1个纹理元素。通常，将原纹理的分辨率按2的幂次方不断地缩小，形成纹理尺寸为1×1、2×2、4×4、8×8…的纹理金字塔结构，如图22-7所示。当多边形距离视点较近时，贴较高分辨率（尺寸较小）的纹理；较远时，贴较低分

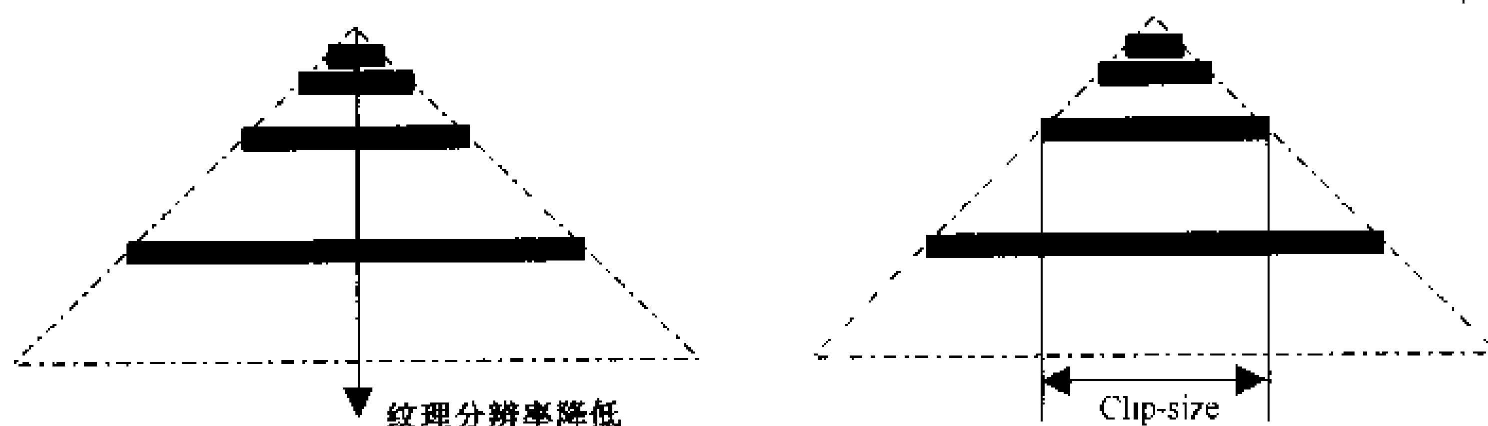


图 22-7 Mip-map 法与 Clip-map 法的异同

辨率（尺寸较大）的纹理。

朱庆（1998）基于“层-块-行列”的思想，提出用 Mip-map 结构来组织多尺度数据，用分块阵列来组织同一层次的数据，而在不同层之间建立相互索引机制，并在 Geostar 的 Windows 与 Windows NT 版本中技术实现。

2. Clip-Map 算法

Mip-map 技术对纹理大小有一定限制，而且必须对影像大小事先做一定的预处理。为了改变这一限制条件，SGI 等公司在软件和硬件设备上做了一些改进，提出了 Clip-map 技术。该技术突破了 Mip-map 技术对纹理大小的限制，允许应用程序使用大数据量的纹理数据。该技术的关键思想是：通过定义一个 Clip-size 来限制不同层的 Mip-map 层的大小，并且在每一时刻只保留必要的一小部分纹理在内存中供进行纹理映射使用。即通过对某一层的 Mip-map 的大小进行判断：若大于 Clip-size，则在映射时使用 Clip-size 进行纹理数据的装载和映射；否则全部装入。Clip-map 与 Mip-map 的区别见图 22-7。

22.2.2 “分治”算法

为了解决数字城市 3D 可视化中的地形范围大、建筑纹理复杂等影响可视化速度的问题，必须设计一种合理的数据结构来存储和管理纹理数据。考虑到纹理分辨率的视点相关性，可以利用“分治”思想，将地形区域分为若干子块（OPENGL 渲染机限定的最小单元为 64 像素×64 像素），然后利用 Mip-map 技术分别建立各子块的多分辨率纹理金字塔模型；在可视化过程中，利用 Clip-map 技术进行建筑纹理的动态映射。

杨必胜（2002）指出：基于“分治”的多分辨率地形影像模型建立可以分为地形原始影像分割和子块无缝拼接（主要指颜色空间）两个阶段；而多分辨率建筑纹理模型的建立则是在地形原始影像模型分割之后，再进行各子块的多分辨率建筑纹理模型构建。多分辨率模型建立的步骤如下：

第一步：读入地形原始影像并判断其大小；

第二步：根据所设定的最小影像单元的大小，将地形原始影像分割为 n 块；

第三步：根据地形原始影像及其3D 表面范围，确定各子块的几何范围与影像数据；

第四步：根据所设定的分辨率级数，建立各子块的多分辨率地形影像金字塔模型；

第五步：按设定的数据结构，存储各子块的地形影像信息，包括影像属性、空间位置、分辨率级别、影像大小和影像数据；

第六步：按第四、第五步原理，完成多分辨率建筑纹理金字塔模型构建和数据库存储；

第七步：结束。

§22.3 纹理映射算法

纹理可区分为2D纹理、3D纹理和几何纹理（张英杰等，1994）。其中，3D纹理和几何纹理是以过程定义的形式来存储和管理的，而2D纹理则是以数据文件或影像文件的形式来存储和管理的。GIS可视化过程的地形影像属于2D纹理的范畴，单一建筑表面的纹理影像也属于2D纹理的范畴。基于"分治"思想建立了地形影像的多分辨率金字塔模型之后，在可视化过程中如何根据视点位置和视线方向，动态地确定影像纹理的合理分辨率，即如何实现多分辨率纹理模型与几何模型之间的映射，是本节要讨论的问题。

22.3.1 纹理映射原理与算法

GIS可视化时，需要通过空间索引、影像装入和纹理映射3个步骤来实现地形影像纹理的表现。地形影像纹理映射的实质是2D的颜色矩阵（u，v）与3D的几何区域（x，y，z）的对应。

1. 纹理映射原理

如图22-8所示，物像空间上某一几何区域或景物的尺寸为h，离视点的距离为d，

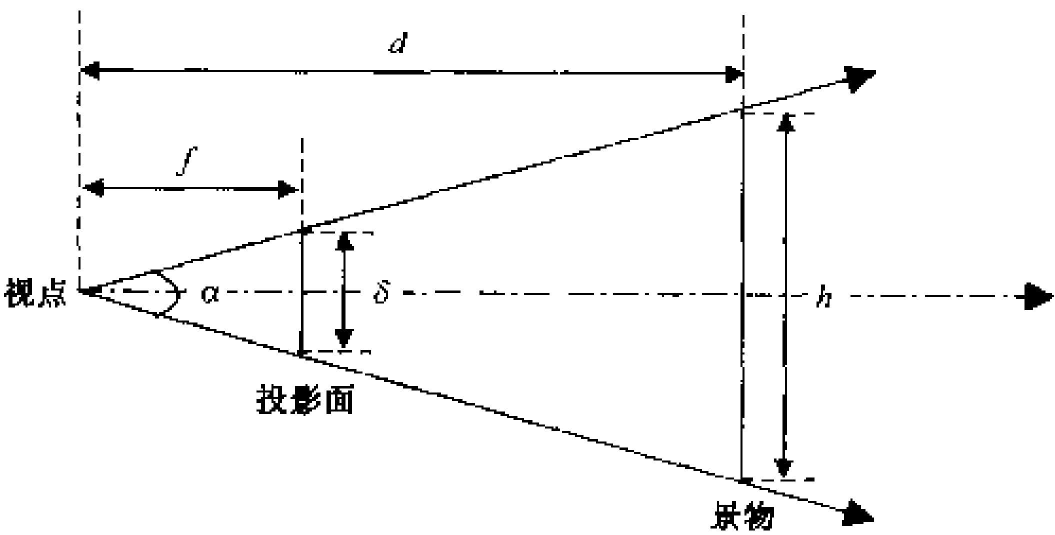


图22-8 景物、投影面与视点的关系

投影面离视点的距离为f。则区域或景物投影到投影面上的尺寸δ为：

$$\delta = \frac{f}{d}h \tag{22-2}$$

若投影面就是计算机的屏幕，且大小比例为1:1，则δ的实质为区域或景物投影到屏幕上占用的像素数目（等于δ除以单位像素的宽度）。反之，取δ为一个单位像素δ_0，则h_0为离视点距离为d的位置的区域或景物的尺寸，h_0为：

$$h_0 = \frac{d}{f}\delta_0 \tag{22-3}$$

区域或景物的几何外形在屏幕上成像之后，就需要将其纹理映射到屏幕上的对应位置。不同分辨率的同一纹理对同一区域进行映射时，纹理的单位像素所对应的实际尺寸是不一样的。此时，由于纹理分辨率的不同，以及屏幕分辨率与纹理分辨率的不同，屏幕上一个像素可能相当于纹理的一个或多个像素。

2. 纹理裁剪算法

地形的几何描述是基于TIN或Grid，地面建筑的轮廓描述也是基于TIN形成多面体。因此，纹理映射的实质上是将纹理分别映射到由TIN或Grid组成的多边形区域上。当多边形区域边界与纹理的图案边界不具备一一对应关系时，需要用多边形区域边界对纹理图案进行裁剪。裁剪算法如下（据张英杰等，1994）：

第一步：将多边形的一个顶点平移到坐标原点；

第二步：将多边形的某一条边旋转变换到与x轴重合；

第三步：将多边形绕x轴旋转，使其变换到$z=0$的平面上；

第四步：计算变换后多边形的矩形包围窗及其窗坐标（x_{max}，x_{min}，y_{max}，y_{min}）；

第五步：确定从矩形包围窗到纹理空间的映射函数：

$$s = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad t = \frac{y - y_{min}}{y_{max} - y_{min}} (0 \leqslant s, t \leqslant 1) \tag{22-4}$$

第六步：将纹理空间变换为单位空间：$s'=0\sim1$，$t'=0\sim1$（s'平行于x轴，t'平行于y轴）；

第七步：根据映射函数，令（s，t）与（s'，t'）形成一一对应关系，对多边形上的点逐个进行纹理采样，完成多边形对纹理图案的裁剪。

3. 色彩混合方式

纹理映射过程中，纹理色彩与被映射的区域表面色彩有3种不同的混合方式：①纹理不透明：当纹理映射到区域表面后，区域的材质色彩全部被纹理色彩所覆盖；②纹理半透明：当纹理映射到区域表面后，仍可看出区域的材质色彩；③混合色彩：当纹理映射到区域表面后，纹理色彩与区域的材质色彩以某种方式（如插值等）进行混合，形成特殊的纹理贴图效果。

22.3.2 纹理分辨率选择算法

为实现多分辨率纹理模型与区域或景物几何对象之间的正确映射，纹理分辨率选择是关键。杨必胜（2002）设计了相应算法来实现多分辨率的纹理分辨率选择与映射。该算法的主要步骤如下：

第一步：根据被映射区域与纹理的映射关系，计算不同分辨率 δ_i 的纹理影像的单位像素所对应的区域长度 h_i，如 n 个不同分辨率 $\{\delta_1, \delta_2, \cdots, \delta_n\}$ 的纹理的单位像素对应 n 个区域长度 $\{h_1, h_2, \cdots, h_n\}$；

第二步：根据纹理的当前分辨率 δ_i 和其单位像素对应的区域长度 h_i，参考当前的视点位置和映射位置的透视投影参数，按公式（22-2）计算出 h_i 在屏幕上投影的大小（所占屏幕像素的数目）：$\zeta_i = \dfrac{f}{d}h_i$；

第三步：根据所设定的像素限定阈值 τ，即单位纹理像素所对应的区域长度在屏幕上投影的大小允许范围 ζ 为：$\zeta_i - \tau \leqslant \zeta \leqslant \zeta_i + \tau$，参考当前的视点位置和映射位置的透视投影参数，按公式（22-2）计算出可选择的纹理的单位像素所对应的区域长度为：$\dfrac{\zeta_i - \tau}{f}d \leqslant h \leqslant \dfrac{\zeta_i + \tau}{f}d$；

第四步：以 $|h - h_i| = \min$，$(i = 1 \sim n)$ 为原则，从所有候选的不同分辨率的纹理中选择出所需要的纹理影像；

第五步：重复第二步至第四步，确定映射区域内不同位置的纹理映射的合理分辨率；

第六步：结束。

§22.4 光相关算法

GIS 3D 可视化与动态交互显示可以分为两个层次，一是预先按既定的观察轨迹逐帧计算场景图像并记录在介质上，然后按某一合适的时间间隔连续播放；二是在可视化过程中随观察轨迹的推进，实时计算场景图像并实时显示。前者虽有局限，但实现相对容易，对计算机性能要求较低；后者一般依赖高档图形工作站，可实现对场景的漫游和实时交互控制。不管采用何种方式，要得到逼真的场景可视效果，除前述的纹理贴图之外，与光相关的处理至关重要，包括空间消隐、可见性判断、透明处理和光照模型等，均是与人的视觉认知密切相关的内容。

22.4.1 消隐算法

消隐处理的目的是确保当前不可见的或被遮挡的场景被屏蔽，以提高可视化速度和质量。消隐处理算法有多种，如画家算法、缓冲区算法、扫描线算法及区域采样算法等。以画家算法为例，该算法的实质是按景物离视点的距离来决定成图顺序，距视点远的景物有限绘制，即让后画的景物去遮挡先画的景物所构成的背景。该法需要对所有景物和构图单元按离动态视点的距离进行动态排序，内存开销很大且排序速度影响可视化速度，一般微机难以承受。

通常，消隐处理中有一个深度测试过程。若某一面元 $f^{(i)}$ 的 z 坐标的最大值小于另一面元 $f^{(j)}$ 或线元 $l^{(k)}$ 的 z 坐标的最小值，则面元 $f^{(i)}$ 完全位于面元 $f^{(j)}$ 和线元 $l^{(k)}$ 之后，根本不可能遮挡面元 $f^{(j)}$ 和线元 $l^{(k)}$，因此不必因面元 $f^{(i)}$ 而对面元 $f^{(j)}$ 和线元

$l^{(k)}$作消隐处理。

谢顺平等（1999）针对微机环境下3D地形模型的可视化，充分利用规则DEM格网面元行列排列的有序性和连贯性，提出只对DEM格网的4个角面元的中心点按距离排序的方法，从而节省了大量的深度计算和排序处理时间。

在消隐过程中，随着视点的变换、投影面的改变、景物位置的变换和景物形变等，景物的几何信息将发生变化，但拓扑关系是不变的，即景物表面的面元、线元、点元是固定不变的。

22.4.2 可见性算法

可见性判断的目的是确保当前非可见的场景（包括面元、线元和点源）被屏蔽，以提高可视化速度和质量。可见性测试包括背面剔除和视窗裁剪两步。以DEM的可视化为例，其可见性判断过程为：① 首先按消隐算法确定的显示顺序，对DEM的格网面元的朝向进行测试，即判断面元的外法矢量与视线的夹角 β：若 $90^\circ < \beta \leqslant 180^\circ$，则该面元为不可见；$\beta = 90^\circ$，则该面元与视线平行，定为不可见；否则（即 $0^\circ \leqslant \beta < 90^\circ$，可见；② 然后对可见面元进行视窗可见性测试，判断经透视投影变换的2D面元是否部分或全部落在视窗范围之内：若在，则对其进行光照计算和显示处理；否则，不处理。

22.4.3 光照算法

光照计算的目的是获取景物面元的明暗值，从而形成场景的浓淡渲染图。光照模型应考虑由环境光源综合引起的泛光、穿过物体表面被吸收并重新发射出来的漫反射光、由物体表面光洁度产生的镜面反射光（高光）等效应，最终以不同颜色和不同亮度表现要素的表面光照特性。

1. 传统的光照模型与算法

传统的光照模型基于场景的几何描述来定义光源分布与性质及景物表面材质的反射特性，即通过某种算法来刻划光源、视点、景物形状、材质4者之间的相互关系。传统的光照模型可分为局部光照模型和整体光照模型两种主要类型。

1）局部光照模型

常用的局部光照模型有辐射度模型和Phong模型。

辐射度模型是Coral C.M.等（1984）提出的，该法把组成环境的曲面看成一个封闭系统，并假定封闭系统中的曲面为漫反射面，进而根据能量平衡原理计算每一曲面上的能量，得到被观察点的辐射度。由于考虑了封闭系统中各曲面之间的能量传递，故该算法能从整体上计算出环境的光能分布，生成的图形逼真且与视点无关。算法为：

$$B_i = E_i + p_i \sum_{j=1}^{n} B_j F_{ij} \tag{22-5}$$

式中：B_i 为第 i 个面片的辐射度；E_i 为第 i 个面片直接发射出的能量；p_i 为第 i 个面

片的漫反射率；F_{ij}为形状因子，定义为从第j面片发出的辐射光能中可以到达第i面片的部分。

当环境被离散为n个面片时，则产生n个辐射度方程。一般，预先设定E_i和ρ_i，通过对方程组的求解，得到式中的未知数B_i和F_{ij}。有半立方体法、半球法和半平面法等求解辐射度方程组的方法。

Phong模型算法为：

$$I_c = k_a I_{pa} + \sum [k_d I_{pd}(N_0 \cdot L_0) + k_s I_{ps}(N_0 \cdot H_0)^n] \tag{22-6}$$

式中：I_c为景物的局部光照亮度；k_a为环境反射的比例系数；k_d为漫反射的比例系数；k_s为镜面反射的比例系数（$k_d + k_s = 1$）；I_{pa}为环境反射分量；I_{pd}为光源垂直入射时反射光的亮度；I_{ps}为镜面反射方向上的镜面反射光亮度；N_0为景物面元的外法向量；L_0为光线向量；H_0为沿光线向量与视线向量的角平分线的单位向量；n为镜面反射光的会聚指数；$\sum$为表示对特定光源求和。

与辐射度模型相比，Phong模型参数直观明确、易于调整，且计算量较小。因而，局部光照一般选择Phong模型。

2）整体光照模型

整体光照模型的代表为Whitted模型（Whitted T.，1980），其算法如下：

$$I = I_c + k_s I_s + k_t I_t \tag{22-7}$$

式中：I_s为来自镜面反射方向的环境光亮度；I_t为来自规则透射方向的环境光亮度；k_t为镜面反射的透射系数；其他参数意义同前。

2. 全光函数模型与算法

Adelson E. H. 和 Bergen J.（1991）基于对早期视觉基本元素的寻找提出了一个全光函数（plenoptic function）来描述整个场景的信息。他们认为：整个世界是一个充满稠密光线的空间，每条光线带有不同的能量，仅此而已（不关心场景中的其他组成形式）；穿过空间中某一点的光线的集合在数学上称为光线锥（pencil）。全光函数正是用一个7维的函数来表示空间中任一点在任意时刻、任意波长下的这一簇光线：

$$p_7 = p(V_x, V_y, V_z, \theta, \phi, \lambda, t) \tag{22-8}$$

式中：V_x，V_y，V_z为任一视点V的空间位置坐标；θ、ϕ为从该视点出发的任一视线的球面角；λ为光线的波长；t为任意时刻。

全光函数是一个非常完备的描述，包括了所有的光线组合和动静模式。场景中光照条件的变化均被通过其第7维参数即时间参数来记录；而场景中光照条件的变化可能来自于场景中任意组成成分的变化，也可能是光源属性的变化，或是景物位置、形状、材质的变化。

基于全光函数的定义和特点，图像（包括遥感影像、摄影图像、人眼视像等）可以看成是遥感器、相机或人眼在某一时刻从空间中给定点沿视线方向某一立体角范围收集到的特定波长下的一束光线（若为整个光线锥中的全部光线，则为全景图 panorama）。换言之，场景就是一个7维的全光函数，而图像仅是该7维全光函数的一个2维截面。

通过某些特定时刻采集到的静态图像，就可以生成场景动态变化的可视结果。

由此发展起来一种全新的可视绘制方法——基于图像的可视绘制技术。该绘制技术既没有景物的几何模型，也没有景物表面的双向反射系数（bia-direction reflection factor，BRDF）和光源分布，有的只是场景的基本组成成分：光线。拍摄图像是采集光线的手段，而可视绘制则是光线的重组过程。因而，基于图像的绘制过程实质就是全光函数的采样、重建与重采样过程。

基于图像的绘制（image-based rendering）与基于图像的建模（image-based modeling），同是计算机图形学中的新技术，均为利用真实场景的影像信息来加速场景建模并提高可视化效果的方法。但两者又互有区别，后者是利用计算机视觉方法恢复出场景的几何模型、材质信息和光源属性等场景信息，然后利用传统的图形学方法得到场景中新视点的视图；前者则无需系统重建过程，直接通过对图像的密集采样来得到系统输出的某种预测，即得到某种形式的全光函数的连续表达。两者之间的区别与联系如图 22-9 所示。

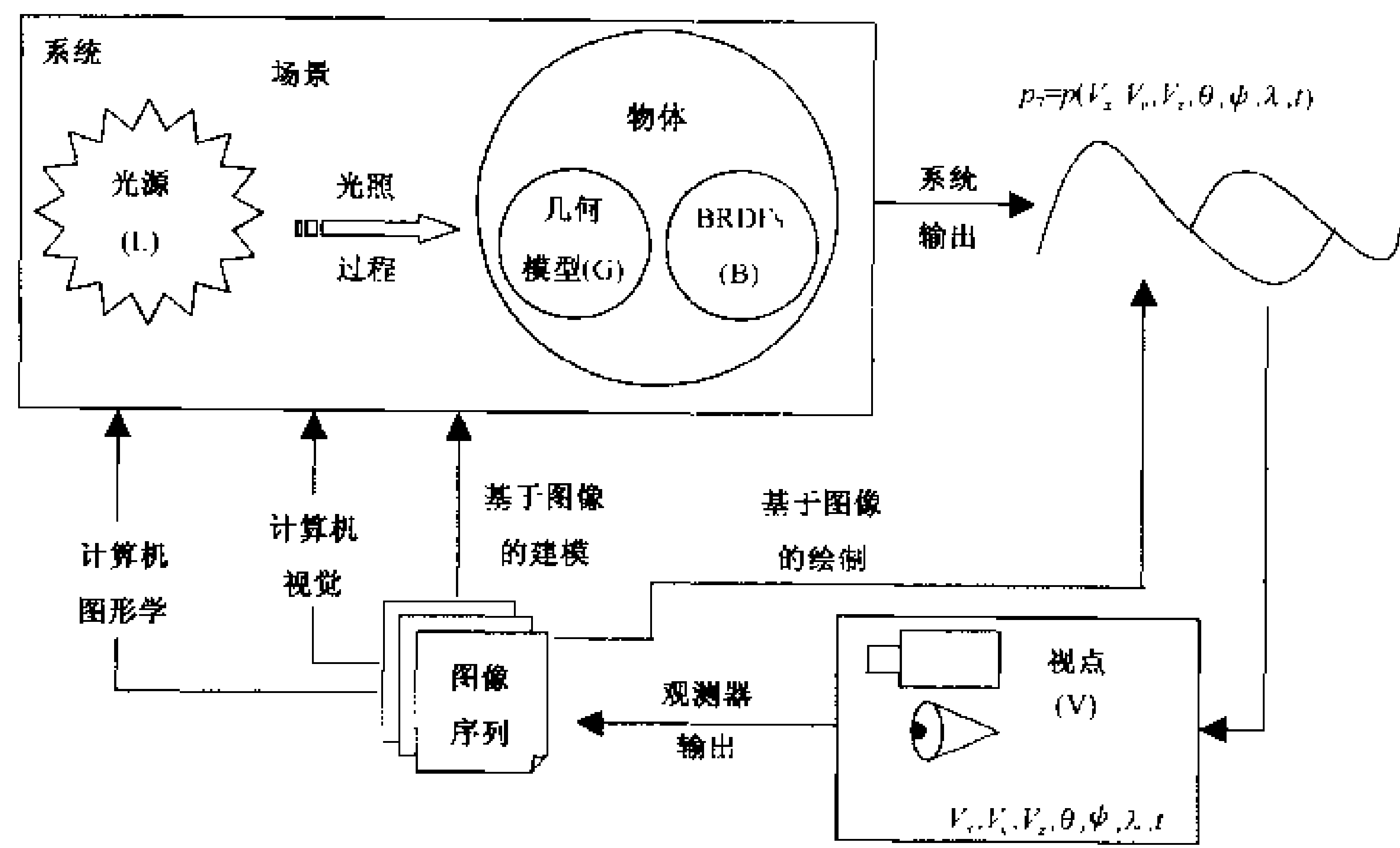


图 22-9 基于图像的统一的可视化系统框架
（沈沉等，2000）

22.4.4 透明算法

某些环境下，需要了解景物内部或景物背后的事物（如地层内部、建筑物背后等），而不是简单地基于可见性判断不显示被遮挡部分的景物。这对于提高可视化效果，以及空间查询与分析能力具有十分重要的作用。

透明处理一般利用光照模型的光线追踪算法进行，包括线性透明、简单非线形透明和折射透明 3 种基本算法。

1. 线性透明算法

线性透明算法由 Newwell（1972）提出，其原理为将前面可见物体表面的亮度与位于透明体后面的物体表面的亮度进行线形组合，形成光强 I：

$$I = I_1 t + (1 - t) I_2 \tag{22-9}$$

式中：I_1 为前面可见面的光强；I_2 为后面紧邻面的光强；t 为前面物体的透射系数（$t=0\sim1$），$t=1$ 表示不透明；$t=0$ 表示完全透明。

显然，当 I_2 所对应的表面也是透明面，则上述算法可以递归进行下去，直到遇到不透明体或背景为止。

该算法虽然简单易行，但存在两点不足：① 未考虑光线在曲面表面的边缘失真现象；② 未考虑光线的路径长度及其在不同介质界面上产生的折射。

2. 简单非线性透明算法

为解决线性透明算法的第一个问题，Kay D. S.（1979）提出基于曲面法向量的 Z 分量的简单非线性近似算法。该算法利用线性插值法首先计算透明系数：

$$t = t_{min} + (t_{max} - t_{min})[1 - (1 - |N_z|^p)] \tag{22-10}$$

式中：t_{max}为前面物体的最大透射系数；t_{min}为前面物体的最小透射系数；N_z 为前面物体表面的单位法矢量的 Z 分量；p 为前面物体的透明幂指数。

显然，经以上处理，由于曲面轮廓处的法矢量的 Z 分量为零，则其透明系数最小，因而透明效果变差。

3. 折射透明算法

本算法是为了解决线形透明算法的第二个问题，包括折射方向计算和透射光强计算两个方面。

1) 折射方向计算

如图 22-10 所示，设光线从空气（折射率为 1）入射到折射率为 n_2 的媒质中，折射光线的单位矢量 $\vec{P}$ 为：

$$\begin{cases} \vec{P} = \vec{D} + \vec{N}\sqrt{1 - |\vec{D}|^2} \\ \vec{D} = \dfrac{1}{n_2}[\vec{I}_n - (\vec{I}_n \cdot \vec{N})\vec{N}] \end{cases} \tag{22-11}$$

式中：$\vec{N}$ 为入射点的法向单位矢量；$\vec{I}_n$ 为入射方向单位矢量；$\vec{D}$ 为按斯奈尔定律求得，其模等于 $\sin\theta'$。

2) 透射光强计算

已知入射点 A 的坐标，透明体的厚度 d 和折射方向角 θ'，可以很容易计算出光线出射点 B 的坐标。令折射光线 $\vec{P}$ 的路径长度为 L，则其穿过整个路径的透明系数为 $t = T \times L$（T 为透明体单位长度的透明系数）。将计算得到的 t 代入式（22-9），即可得到透射光强度。

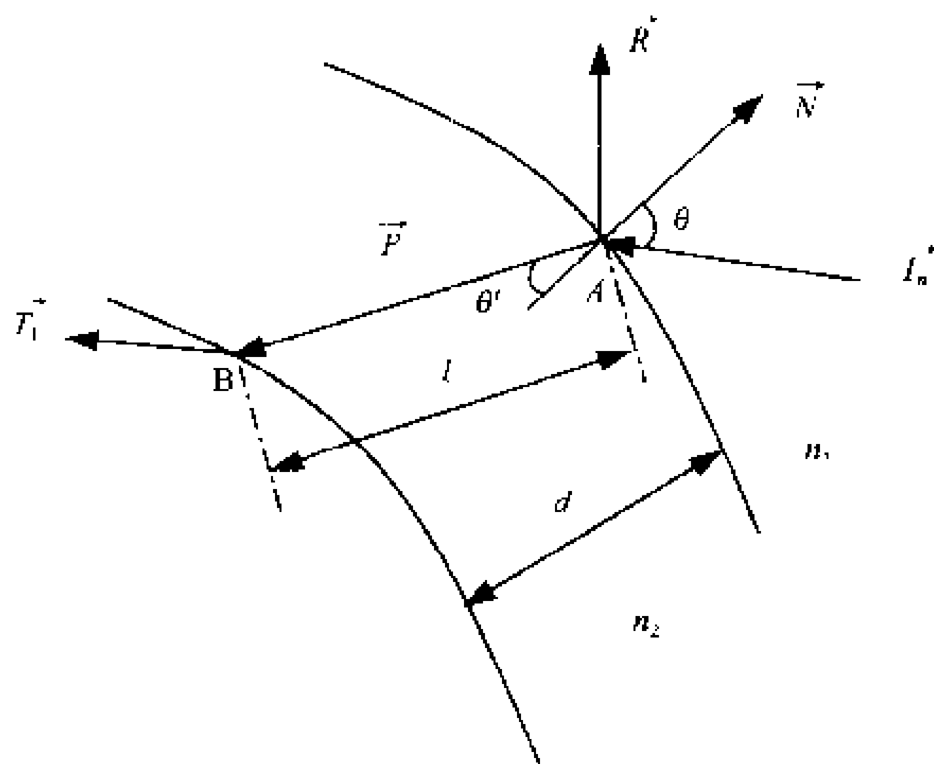


图 22-10 折射算法原理

除以上讨论的空间消隐、可见性判断、透明处理和光照算法之外，在 GIS 可视化过程中与光相关的处理还有倒影处理、阴影处理、天空背景处理（龚明伦，秦开怀，1996）等众多方面，限于篇幅，不再赘述，读者可查阅有关文献资料。

参 考 文 献

高 俊 2000. 地理空间数据可视化. 测绘工程 9 (3): 1－7

龚明伦，秦开怀. 1996. 建筑造型及其环境图像的渲染技术. 中国图像图形学报. 1 (5, 6): 400－406

林 珲，龚建华. 2001. 虚拟地理环境. 高等教育出版社

潘志庚等. 1996. 虚拟环境中多细节层次模型的自动生成算法 软件学报. 7 (9): 526－531

秦开怀，龚明伦. 1996. 基于图像与图形综合的光线跟踪技术 计算机研究与发展. 34 (增): 225－230

沈 沉，沈向洋，马颂德. 2000. 基于图像的光照模型研究综述 计算机学报 23 (12): 1261－1269

孙家广等. 1986. 计算机图形学. 北京: 清华大学出版社

孙 敏，马蔼乃，毛善君. 2001. 3DGIS 中树的表达与可视化研究. 计算机辅助设计与图形学学报. 13 (10): 901－905

吴立新等. 2000，论 21 世纪的矿山——数字矿山. 煤炭学报，25 (4): 337－342

吴立新，殷作如，钟亚平. 2003，再论数字矿山: 特征、框架与关键技术 煤炭学报. 28 (1): 1－7

杨必胜. 2002. 数字城市的三维建模与可视化技术研究 [博士学位论文]. 武汉: 武汉测绘科技大学

张英杰，张铁昌. 1994. 基于光线跟踪算法的纹理映射技术. 计算机工程 20 (5): 30－33

朱 庆. 1998. 3D 动态交互可视化模型: GIS 中的 3D 表示与分析. 武汉测绘科技大学学报 23 (2): 124－127

（美）D. F. 罗杰斯（梁友栋等，译）. 1987. 计算机图像学的算法基础 北京: 科学出版社

Adelson E. H., and Bergen J., 1991. Computational Models of Visual Processing. Cambridge, MIT Press

Chen Z. T., and Guevara J. A., 1987. Systematic selection of very important points (VIP) from digital terrain model for construction triangulation network. Proc. Auto-Carto 8. Baltimore. 50－60

Coral C. M. et al., 1984. Modeling the integration of light between diffuse surface. In: Proc. of SIGGRAPH'84. New York: The Assoc. Computing Machinery, INC. 213－222

De Floriani et al., 1983. A delaunay-based method for surface approximation. In: Eurographics'83. 333－350

De Floriani et al., 1984. A hierarchical data structure for surface approximation. Computer and Graphics. 1984. 8 (2): 475－484

Duchaineau M. et al., ROAMing terrain: real-time optimally adapting meshes. IEEE Visualization'97 Proc. http://www.llnl.gov/graphics/ROAM

Evans W. et al., Right triangular irregulay network. http://www.cs.arizona.edu./people/will/.

Fowler R. J., and Little J. J., 1979. Automatic extraction of irregular network digital terrain models. Computer Graphics. 13 (2): 199~207

Gahengan M., 1999. Four barrier to the development of effective explorary visualization tools for geoscience. Int J Geographical Information Systems. 13 (4): 289-309

Garland M., and Heckbert P. S, 1995. Fast Polygonal Approximation of Terrenes and Height Fields. Tech. Report of CMU-CS-95 181. http://www.cs.cmu.edu/ garland/scape.

Garland M. and Paul S. H., 1998. Simplifying surface with color and texture using quadric erro metrics. Visualization'98

Hoppe H., 1996. Progressive Meshes. Computer Graphics (Proc SIGGRAPH'96) 99-108

Hoppe H., 1998. Smooth view-dependent level of detail control and its application to terrain rendering. IEEE visualization'98. 35~42

Kay D. S., 1979. Transparency for computer synthesized images. Computer Graphics. 13: 158~164

Jonathan Raper edited: Three Dimensional Application in GIS London.

Lawson C. L., 1977. Software for C^1 surface interpolation. Mathematical Software III New York, Academic Press. 161-194

Lee J., 1989. Drop heuristic conversion method for extracting irregular network for digital elevation models. In: GIS/KIS'89 Proc., Volume I. 30~39

Lindstrom P. et al., 1998. Real-time continuous level of detail rendering of height fields. In: proc. IEEE Visualization'98. 279~286

Rippa S, 1992. Long and thin triangles can be good for linear interpolation. SIAM J. Numerical Analysis. 29 (1): 257~270

Robin A. M. and Tome J. M. K., 1989. Visualization of digital terrain models: techniques and applications. In J. Raper ed, three Dimensional Application in GIS, London

Scarlatos L and Pavlidis T., 1992a. Optimizing triangulation by curvature equalization. In: Proc. Visualzation'92. IEEE Computer Soc. Press. 333~339

Scarlatos L. and Pavlidis T., 1992b. Hierarchical triangulation using cartographic coherence CVGIP: Graphical models and Image Processing. 54 (2): 147~161

Schroeder W et al., 1992. Decimation of triangles meshes. Computer Graphics. 26 (2): 65~68

Slocum T. A., 1999. Thematic Cartography and Visualization. Prentice Hall

Whitted T., 1980. An improved illumination model for shaded display CACM. 23: 343~349

Von H. B., 1987. 1987. Accurate triangulation of deformed intersection surface. Computer Graphics. 21 (4): 103-110

Xia J. C. and Varshney A., 1996. Dynamic view-dependent simplification for polygonal models. IEEE Transaction on Visualization and Computer Graphics. 13 (2): 327-334

第二十三章　空间数据挖掘与知识发现算法

数据、信息与知识是当今社会人所共知的术语，内涵却不一样。数据（data）是关于客观事物的属性、数量、位置及其相互关系等的符号（数学符号、制图符号、声音图像等）描述，是信息的载体，同一个数据在不同的场合下可以表示不同的信息；信息（information）则是特定场合下的、有意义的数据，是对数据的解释，同一个信息在不同的场合下可以用不同的数据进行表示；知识（knowledge）既是一个或多个信息关联在一起形成的有应用价值的信息结构，也是不与其他信息相关联的信息（事实）。从数据到信息是一个数据处理的过程，包括查询、简单统计、特征提取等，在数据库管理系统中通过查询和统计功能实现；从信息到知识则是一个认知的过程，当今，特别是信息时代，往往需要人与机器协调完成。三者之间的关系如图 23-1 所示。

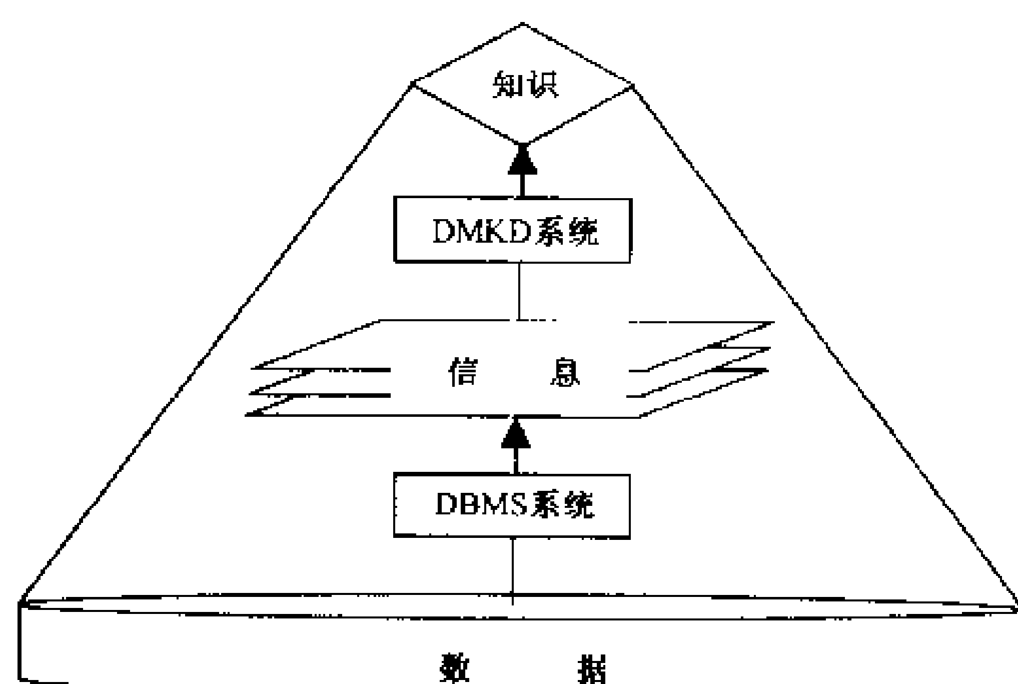


图 23-1　数据、信息与知识的金字塔结构
（据邸凯昌，2001）

数据挖掘（data mining）起源于从数据库中发现知识（knowledge discovery in database，KDD），于 1989 年第十一届国际联合人工智能会议上被定义为："从数据中发现隐含的、先前不知道的潜在有用的信息的非平凡过程"（Frawley W. et al.，1991）。由 KDD 衍生出来的其他术语还有：知识提取（knowledge extraction）、信息发现（information discovery）、信息收获（information havesting）和数据考古（data archaeology）等。为了统一认识，Fayyad U. M. et al.（1996a）对 KDD 和数据挖掘进行了区分和重新定义。KDD 的新定义为：KDD 是从数据中辨别有效的、新颖的、潜在和有用的、最终可理解的模式的过程。数据挖掘的新定义为：数据挖掘是 KDD 中通过特定的算法，

在可接受的计算效率下生成特定模式的一个步骤。可见，重新定义后的 KDD 是一个包括数据选择、预处理、交换、挖掘、评价等步骤，并最终得到知识的过程；而数据挖掘只是其中的一个关键步骤。尽管如此，人们还是经常将两者等同起来，即使是单独使用数据挖掘一词，其实也包括数据预处理和结果评价等内容。此外，人们也趋于将两者一起使用，称为数据挖掘与知识发现（Ddata mining and knowledge discovery，DMKD）。

近年，DMKD 技术作为数据库、人工智能、数据处理和可视化技术的集成，越来越引人注目。已在军事、天文、商业、金融、银行、保险、电信、交通等领域得到广泛应用；DMKD 学术会议也非常活跃，国际上第一本 DMKD 杂志《Data Mining and Knowledge Discovery》已于 1997 年创刊。DMKD 已成为 20 世纪末和 21 世纪初信息技术领域的最热门的课题之一。空间数据挖掘（spatial data mining，SDM）是 KDD 在智能化 GIS 中的具体应用（knowledge discovery from GIS）（Li D. R. et al.，1994；李德仁，1995），是 DM 的一个新的研究分支，其实质是从空间数据库中挖掘、发现时空系统中内在的、有价值的信息、规律和知识的过程，包括空间模式与特征、空间与非空间数据之间的概要关系等（Lu W. et al.，1993）。本章将重点讨论空间数据挖掘与知识发现（spatial data mining and knowledge discovery，SDMKD）的有关内容和算法，介绍 SDMKD 在卫星遥感气象规律研究中的应用进展，并以矿山为例进行 SDMKD 的系统分析与设计。

§23.1 空间数据仓库与数据挖掘

23.1.1 空间数据仓库技术

数据仓库（data warehouse）是与数据挖掘密切相关的一项热点技术，它是由传统数据库发展而来的一个综合的、面向分析的数据库环境。数据仓库之父 Inmon W. H. 给数据仓库下的定义是：数据仓库是集成的、面向主题的、随时间而变化的、持久的数据集合，用于决策管理中的决策制定过程。它的基本思想是事先把应用所需的所有数据从大量的运作数据库中抽取出来，并转化成一定的格式，以加快查询和分析的速度。数据仓库和数据库的不同在于：数据库是一种通用的平台，用来管理企业的数据；而数据仓库则主要是一种概念，在此概念下进行的构造过程称为数据仓库处理。

随着数据仓库技术的发展，如今数据仓库的实施策略已从原先的“自顶向下”模式发展到了“自底向上”、“平行开发”、“有反馈的自顶向下”、“有反馈的自底向上”和“有反馈的平行开发”等模式。其中前 3 者都只在建设数据集或数据仓库的过程中考虑了用户需求，而没有考虑如何将用户的反馈信息不断地反应到数据集市和数据仓库的建设中，因而不能适应信息内容动态变化；后 3 者则是针对这些问题的合理解决方案。

空间数据仓库（spatial data warehouse）则是在数据仓库技术基础上引入空间数据而发展起来的一类特殊的数据仓库。它是根据主题截取从瞬态到区段的不同时空尺度上的信息，从而为各种领域的研究和政策决策提供信息服务的系统。空间数据仓库不是仅仅只在数据仓库上加一层空间的外衣，以地图的形式来反映某些结果，而是真正地以空间数据库为基础，能够进行复杂的空间分析，反映自然界不同时空尺度下的动态变化趋

势，为决策者提供及时、准确的信息。空间数据仓库的主要特征如下：

（1）空间数据仓库是以空间位置数据为标志的；

（2）空间数据仓库是面向主题的，每一个主题大体对应一个宏观的分析领域；

（3）空间数据仓库是集成的，它以各种面向应用的数据库系统为基础，通过元数据刻画的抽取和聚集规则将它们集成起来，从中得到各种有价值的数据。

（4）空间数据仓库可以管理时间序列的历史数据，数据有时间属性，能反映自然界的时间变化趋势。

由于空间数据结构的复杂性，空间数据仓库技术的发展要落后于商务数据仓库的发展。对于数据规模不是很大的空间数据，其空间数据仓库可以在现有的某一GIS系统中开发，因而可以看成该GIS的一个特定的应用系统；对于数据较大、数据类型多样的空间数据，则应基于空间数据引擎（如ESRI的SDE，MapInfo的Spatialware等）进行开发空间数据仓库，其特点是借用通用性关系型数据库系统来管理空间数据，数据安全性和一致性方面技术成熟，便于维护。

由于数据仓库中的数据是经过选择、整理和集成的，为数据挖掘提供了良好的数据基础。因此，在数据仓库中进行数据挖掘往往要比在原始数据库中进行数据挖掘更加有效。传统数据库的主要数据处理与分析技术是在线分析处理（on line analytical pocessing，OLAP），其主要功能是多维数据分析与报表生成。OLAP可以看成是一种简单的数据挖掘技术，而归纳（generalization）、关联（association）、分类（classification）、聚类（clustering）、趋势分析（trend analysis）等则是DMKD的深化技术，这些技术之间的相互结合可以提高DMKD的效率。

23.1.2 可发现的知识类型

1. 一般数据库中的知识

一般认为，可以从数据仓库发现以下5类知识：

（1）概括型知识（generalization）：根据数据的微观特性发现其表征的带有普遍性的和较高层次的中观、宏观性知识，通常是对数据的概括、精练和抽象；

（2）分类型知识（classification & clustering）：反映同类事物共同性质的特征型知识和不同事物之间差异型的特征知识，通常是对数据的汇聚或属性分类；

（3）关联型知识（association）：反映一个事件和其他事件之间相互依存或相互关联的知识，又称依赖（dependency）关系，通常用于数据库中的归一化和查询优化；

（4）预测型知识（prediction）：通过时间序列模型，由历史的和当前的状态去预测未来可能出现的情况，它实质上是一种以时间为关键属性的关联知识；

（5）偏差型知识（deviation）：是对数据差异和极端情形的描述，如标准类之外的特例，数据聚类之外的离群值，实际观测值和系统预测值之间的显著差等。

2. 空间数据仓库中的知识

基于GIS与遥感应用，一般认为（Li D R. et al.，1994，1997；邸凯昌等，1996，1997；Han J.，1996；Koperski K. et al.，1996），可以从空间数据仓库中发现的知识有

以下8种：

1）普遍的几何知识（general geometric knowledge）

所谓普遍的几何知识，是指关于目标的数量、大小、形态特征等的普遍性知识，如点状目标的位置、大小等，线状目标的长度、大小和方向等，面状目标的周长、面积、几何中心等。可以通过计算或统计得出GIS中空间目标某种几何特征量的最小值、最大值、均值、方差、中数等，还可以统计出有关特征量的直方图等。

2）空间分布规律（spatial distribution regularities）

空间分布规律是指目标在地理空间中垂直向、水平向以及垂直与水平联合的分布规律。垂直分布是指空间目标沿地理高程的分布，如植被覆盖与生物多样性的海拔变化，作物生长的坡度变化等；水平分布是指地物沿地理区域的平面分布，如水稻亩产的区域分异性，基础设施的城乡差异性等；垂直与水平的联合分布是指目标在高程和区域方面的同时变化，如风土人情的流域差异性、气候特征的地理分异性等。

3）空间关联规则（spatial association rules）

空间关联规则是指空间目标的相邻、相连、共生、包含等关系，如道路与河流的相连，国家与国家的相邻等。

4）空间分类（聚类）规则：（spatial classification/clustering rules）

空间分类规则是根据目标的空间或非空间特征进行类别划分的规则；而空间聚类规则是根据目标的聚散程度进行类别划分的规则，可用于GIS的空间概括和空间综合。空间分类和空间聚类的共同之处是：都是对目标空间的再划分，划分的标准是类内差别最小而类间差别最大；不同之处是：分类是有导师的（supervised）而聚类是无导师的，即分类是事先知道类别数和各类的典型特征，而聚类则事先不知道。

以遥感图像分析为例，一般性遥感图像处理是针对一幅图像或一个试验区的多幅图像进行处理和分析，找出某种结论性的东西；而遥感数据挖掘则强调对大量数据的处理、分析和对比，进而找出共性和特性，总结出规律和规则，而这些规律和规则在后续的图像分析中具有指导作用。

5）空间特征规则（spatial characteristic rules）

空间特征规则是指某类或某几类空间目标的几何与属性的普遍特征，是对共性的描述。比如："高速公路一般比较直"、"高速公路一般6车道"是两条描述高速公路普遍特征的空间特征规则。普遍的几何知识属于空间特征规则的一类，将它分离出来单独作为一类知识是由于它在遥感影像解译中有十分重要的作用。

6）空间区分规则（spatial discriminate rules）

空间区分规则特征是指两类或多类目标间几何或属性的不同特征，是对个性的描述。"高速公路一般比较直，高等级公路一般稍弯一些"、"高速公路一般6车道，高等

级公路一般 4 车道”是两条描绘高速公路与高等级公路特征差别的区分规则。它不同于空间分类规则：分类规则是对空间目标进行明确分类，强调的是分类精度，为了保持分类精度，一般在较低的层次进行分类；而区分规则是对已知类别对象的对比，一般是较高层次上的描述。

7）空间演化规则（spatial evolution rules）

空间演化规则是指空间目标的几何、属性特征随时间而变化的规律。这一规律的发现必须基于时空数据库或同一区域的多个时相的数据。

8）面向对象的知识（object oriented knowledge）

它是关于某类复杂对象的子类构成及其普遍特征的知识。比如对于一个小区对象，由住宅楼、服务楼、幼儿园、绿地、健身区等子类对象构成，相互之间的空间穿插和均衡布局等构成该小区的总体特征。

23.1.3 SDMKD 在 GIS 与遥感中的应用

基于 SDB 的 SDBM 系统可以大致分为 3 层结构，如图 23-2 所示。SDMKD 在 GIS 与遥感中有广泛的应用，主要包括 GIS 智能化分析、遥感影像解译和空间决策支持。

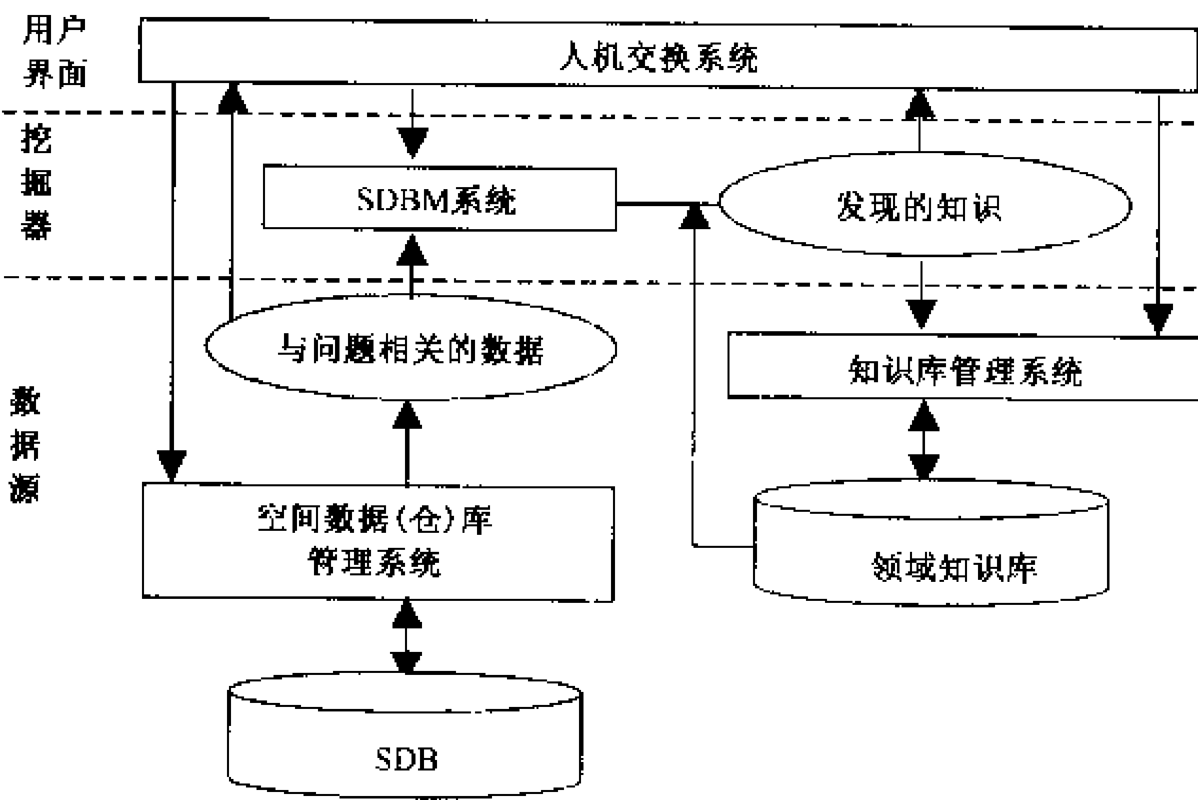


图 23-2 空间数据挖掘的体系结构
（据周海燕等，2002）

1. GIS 智能分析

现有的 GIS 系统一般已经具有强大的空间数据管理、制图、查询、统计和空间分析功能，但缺乏知识的表达、获取和应用机制。因此，现有 GIS 的总体智能化水平还较低。SDMKD 技术恰好可以弥补这一不足，因为 SDMKD 所获取的知识与现有 GIS 分

析工具获得的信息相比，更加概括和精练，并可发现现有 GIS 分析工具无法获取的隐含的模式和规律，可以具备与人类思维相近似的推理模式。

2. 遥感影像解译

在遥感影像自动解译、目标识别和特征提取过程中，利用影像以外的信息提高工作效率和工作质量是遥感技术的一项重要议题。如果通过 SDMKD 技术，从已有的 GIS 和 RS 资料库中发现知识，就可以用于遥感自动解译、目标识别和特征提取中的约束、辅助和引导过程，解决同谱异物、异谱同物的问题，从而减少分类识别的疑义度，提高解译的可靠性、精度和速度。可以预见，SDMKD 技术定将促进遥感与 GIS 的进一步结合和智能化集成。

3. 空间决策支持

现有 GIS 系统虽然已经开始与各类专家系统（experts system，ES）结合起来，包括用来解决电信规划方面（苏辉等，2001）等方面的问题。但总的来说，现有的 GIS 与 ES 结合虽然可以在较大程度上解决知识的表达与应用问题，但在知识获取方面依然存在一大“瓶颈”。如果在 GIS 中引入了 SDMKD 技术，就有可能自动、半自动地从大量的空间数据中发现一些特定的知识或普遍规律，来辅助决策过程，提高复杂系统决策的科学性、合理性和决策效率。

§23.2　算法分类剖析

由于 SDMKD 是一类典型的交叉学科的新领域，它综合了测绘与遥感科学、地理学、数据库、专家系统、模式识别、数理统计、模糊数学、人工智能、图形图像学、可视化等领域的有关技术，因而 SDMKD 算法丰富多彩。一般地，可以基于数据库种类、知识种类和所使用的技术 3 个方面来对 SDMKD 方法进行分类（李永平等，1998）；汪闽等（2002）从知识类别入手，将近年来出现的主要 SDMKD 方法分为以空间信息泛化为目标、以空间关联规则发现为目标和以空间分类（聚类）为目标的 3 类 SDM 技术；邸凯昌（2001）则针对空间数据的特点，在国内外有关研究的基础上总结提出了 14 种代表性的 SDMKD 方法。我们根据各类方法本身的数学基础和现有应用背景，将 SDMKD 算法进行如表 23-1 所示的归类。

表 23-1　SDMKD 算法分类

概率统计类	模糊数学类	人工智能类	图形图像类	空间分析类
统计法	模糊集法	诱导学习法	图像分析与模式识别	拓扑结构分析法
归纳学习法	云理论	神经网络	可视化	空间缓冲分析法
探测分析法		遗传基因法		空间距离分析法
聚类法		Rough 集法		空间密度分析法
关联规则法		Rough 集 + 云理论		空间叠置分析法
证据理论				地形分析法
				网络分析法

23.2.1 概率统计类算法

1. 统计法（statistical approach）

统计法是数字型空间数据分析的常用方法，包括多元统计中的回归分析、方差分析、主成分分析、因子分析等。通常，统计法要求数据之间满足统计不相关假设。但实际上，由于空间对象的属性经常受邻近对象的影响，相关性很强，这就限制了统计法在 SDMKD 中的应用。以变异函数（variogram）和 Kriging 法为代表的地质统计学（Geostatistics）方法，是地学领域特有的统计分析方法。该法针对地学尤其是地质、矿体现象的高度相关性，考虑了空间数据的相关性，因而更加合理，可以在 SDMKD 中发挥作用。

2. 归纳学习法（inductive approach）

归纳学习法是从大量的经验数据中归纳抽取一般的规则和模式。归纳学习的算法很多，大部分来源于机器学习领域。主要算法有（周海燕等，2002）：Michaski 等的 AQ11、AQ15，洪家荣等的 AE1、AE9，Hunt 的 CLS，Quinlan 的 ID3、C5.0 等，其中最著名的是 C5.0。

3. 探测分析法（exploratory data analysis，EDA）

EDA 是相对于传统的统计分析而言的，它不预设数据具有某种分布或规律，而是完全从无知入手，一步一步地、试探性地分析数据，逐步地理解和认识数据。EDA 将统计分析与可视化技术结合起来，即采用动态统计图形和动态链接窗口技术将数据及其统计特征显示出来，可以随时发现数据中非直观的数据特征及异常，因而提高了数据分析的灵活性和深度，特别适合于 GIS 的数据分析。EDA 技术在 SDMKD 中还用于选取感兴趣的数据子集，即进行数据聚焦（data focusing），可初步发现隐含在数据中的某些特征和规律。

4. 聚类法（clustering approach）

聚类法是按一定的距离或相似性测度将数据分成一系列相互区分的组。它与诱导学习法的区别在于不需要背景知识而直接发现一些有意义的结构与模式。经典的聚类法包括 K-mean、K-meriod、ISODATA 等。经典的基于统计学的聚类分析方法对大数据量属性数据库操作时存在速度慢、效率低等问题。近年来，围绕 DMKD 领域，发展了 CLARANS（Ng R. et al.，1994）、BIRCH（Zhang T. et al.，1996）、DBSCAN（Ester M. et al.，1996）、Murray（Murray A. J. et al.，1998）等算法。此外，在 SDMKD 领域，有必要针对图形数据库发展新型的空间聚类方法。

5. 关联规则法（association rule approach）

关联规则法首先由 Agrawal R. 等（1993）针对超市商品销售搭配问题的研究而提出。包括单层次关联规则法、多层次关联规则法两种基本模式。

Agrawal R. 提出的 Apriori 算法是单层次关联规则法的代表，其基本思想为：先统计出多种商品在一次购买中共同出现（称为项目集）的频次；然后将出现频次多（称为大项目集）的商品种类的搭配转换为关联规则。该算法的核心是用前一次扫描数据库的结果来产生本次扫描的候选项目集，从而提高搜索的效率。算法详细过程见 Agrawal R. 等（1994）。Park J. S. 等（1995a）提出的 DHP 算法应用"hash"技术进一步减少候选项目集的数量，更有效地提高了速度。

从事务数据库中挖掘出关联规则，可以看做是布尔逻辑运算过程，因为事务中仅包含二值属性（项目）。为了进一步从大型关系型数据库中挖掘定量关联问题，Srikant R. 等（1996）把定量关联规则问题映射为布尔关联规则问题。如果定量属性的定义域很大，则首先把属性值划分到不同的区间中，然后把每一个属性-区间对映射为布尔属性。经过映射之后，任何一种挖掘布尔关联规则的算法均可用于挖掘定量关联规则。

Han J. 等（1995，2001）、Srikant R. 等（1995）还对单层次关联规则挖掘算法进行了扩展，研究了在多层次结构中挖掘关联规则的算法。Park J. S. 等（1995b）和 Cheung D. W. 等（1996），KoperskiK. 等（1995），以及 Kamber M. 等（1997）则分别报道了关联规则并行挖掘算法、空间关联规则的挖掘和元规则引导的多维关联规则挖掘问题。

6. 证据理论（evidence theory）

证据理论又称 Dempster-Schafer 理论，是经典概率论的一种扩展方式。该理论首先由 Dempster 于 20 世纪 60 年代提出，在 70 年代中期由 Schafer 进一步发展，形成处理不确定信息的证据理论，其特色就是划清了不确定与不知的界限。

在证据理论中，一个样本空间称为一个识别框架（frame of discriminate），用 Ω 表示。Ω 由一系列对象（元素）构成，对象之间两两互斥，并且包含当前要识别的全体对象。Ω 的所有子集的集合记为 2^{Ω}，若 Ω 中有 n 个对象，则有 2^{Ω} 个子集。每个子集对应一个命题（证据或结论）。证据理论的基本问题是：已知识别框架 Ω，判明 Ω 中一个先验的未定位元素属于 Ω 中某个子集的程度。

Anand S. S. 等（1996）提出了一个基于证据理论的通用数据挖掘架构 EDM，该架构由两大部分组成：一是表达数据和知识的数据 mass 函数和规则 mass 函数，这些函数是证据理论中的 mass 函数的推广；二是用于发现知识的 mass 函数算子。该架构的运作思想是：首先将关系数据模型映射到 EDM 框架中，关系模型中的关系对应 EDM 中的搜索空间，关系数据模型中的空属性值 NULL 在 EDM 中作为"未知"来处理。经过映射，从数据库发现知识的问题就变为以数据库中的证据为基础，对于前件识别框架中选择特定的规则前件值，从后件识别框架中选择合适的后件值。因而，数据库中知识发现的过程就变成了对数据 mass 函数和规则 mass 函数的一系列运算，最终的规则 mass 函数可表达为"If…，then….”规则。目前，EDM 已开发的算子有组合算子（combination operator）、归纳算子（induction operator）、领域算子（domain operator）等。在 EDM 框架下，DMKD 的技巧主要是根据任务来选择或开发不同的算子。

总之，EDM 框架使得知识发现的过程统一地表现为对数据和规则的 mass 函数运算，可以方便地处理空属性或丢失属性值，并且可以在知识发现过程中引入领域知识。

同时，EDM 算法本质上是并行的，从而为利用并行处理技术来实现对海量、异构及分布式空间数据库的 SDMKD 提供了基础，应用前景乐观。

23.2.2 模糊数学类算法

1．模糊集理论（fuzzy sets theory）

Zadeh L. A.（1965）扩展经典集合理论，提出了模糊集理论，用于处理自然界和人类社会中的模糊现象与相关问题。模糊集理论的核心是用隶属度来刻划模糊事务的亦此亦彼性。模糊集理论在遥感图像的模糊分类、GIS 模糊查询、空间数据不确定性表达和处理方面也得到大量应用。在 SDMKD 中，模糊集理论也得到了应用，如用语言值概括数据和进行模糊知识表达等（Yager P. R.，1991；Ziarko W.，1994）。

尽管如此，模糊集理论有一个固有缺陷，即一旦用一个精确的隶属函数来描述模糊集，模糊概念就被强行纳入到精确数学的范畴，此后在概念的定义、定理的叙述及证明等数学思维环节中，就不再有丝毫的模糊性了。这在一定程度上限制了模糊集理论在 SDMKD 中应用的范围和程度。

2．云理论（cloud theory）

针对模糊集理论的固有缺陷，李德毅院士提出了定性定量不确定性转换模型——云模型（cloud model），进而把定性概念的模糊性和随机性结合起来，构成定性定量之间的相互映射，作为知识表达的基础（李德毅等，1995）。随后，又提出了虚拟云（virtual cloud）的概念及构造虚拟云的方法，提出了云变换（cloud transform）和不确定性推理（reasoning under uncertainty）的思想（Li D. Y. et al.，1997，1998）。其中，虚拟云和云变换用于概念层次结构生成和概念提升；云推理用于不确定性预测等。邸凯昌（2001）在此基础上，对一维正态云模型进行了扩展，研究提出了二维和多维正态云模型，以适应空间数据库的需要；并将正态云模型扩展至Γ云、三角云、梯形云等多种形状的云模型，进而形成了一个较完整的云模型系列。

以一维云模型为例，设 U 是一个论域 $U=\{x\}$，T 是与 U 相联系的语言值。U 中的元素 x 对于 T 所表达的定性概念的隶属度 $C_T(x)$（或称 x 与 T 的相容度）是一个具有稳定倾向的随机数，隶属度在论域上的分布称为隶属云，简称云。因此，“云”实质上就是用语言描述的某个定性概念与其数值表示之间的不确定性转换模型。$C_T(x)$ 在 [0，1] 中取值，云是从论域 U 到区间 [0，1] 的映射，即：

$$\begin{aligned} &C_T(x): U \rightarrow [0,1] \\ &\forall x \in U, x \rightarrow C_T(x) \end{aligned} \tag{23-1}$$

一维云模型有以下特点：

（1）论域中所有 x 到区间 [0，1] 的映射是一对多的转换，即 x 对于 T 的隶属度是一个概率分布而非固定值；换句话说，云是由云滴组成的，每个云滴是定性概念在数量上的一次具体实现；

（2）云的总体是有边界的，从模糊集理论的观点来看，云的数学期望曲线（mathematical expected curve，MEC）是其隶属曲线；

（3）云的厚度是不均匀的，反映了隶属度的随机性大小；靠近概念中心或远离概念中心处隶属度的随机性较小，所以云的顶部和底部最薄；而离概念中心不近不远处的隶属度的随机性较大，所以云的腰部最厚。

如图 23-3 所示，一维云的数学特征可以用期望值 Ex（expected value）、熵 En（entropy）和超熵 He（hyper entropy）来表征。期望值 Ex 是该定性概念在论域中的中心值，100%隶属于该定性概念；熵 En 是定性概念模糊度的度量，反映了在该论域中可被该定性概念接受的数值范围，熵越大，表示该概念所接受的定量数值范围越宽，概念越模糊；超熵 He 是关于熵的熵，反映云滴的离散程度，超熵越大，云滴的离散程度越大，隶属度的随机性越大，云的厚度也越大。

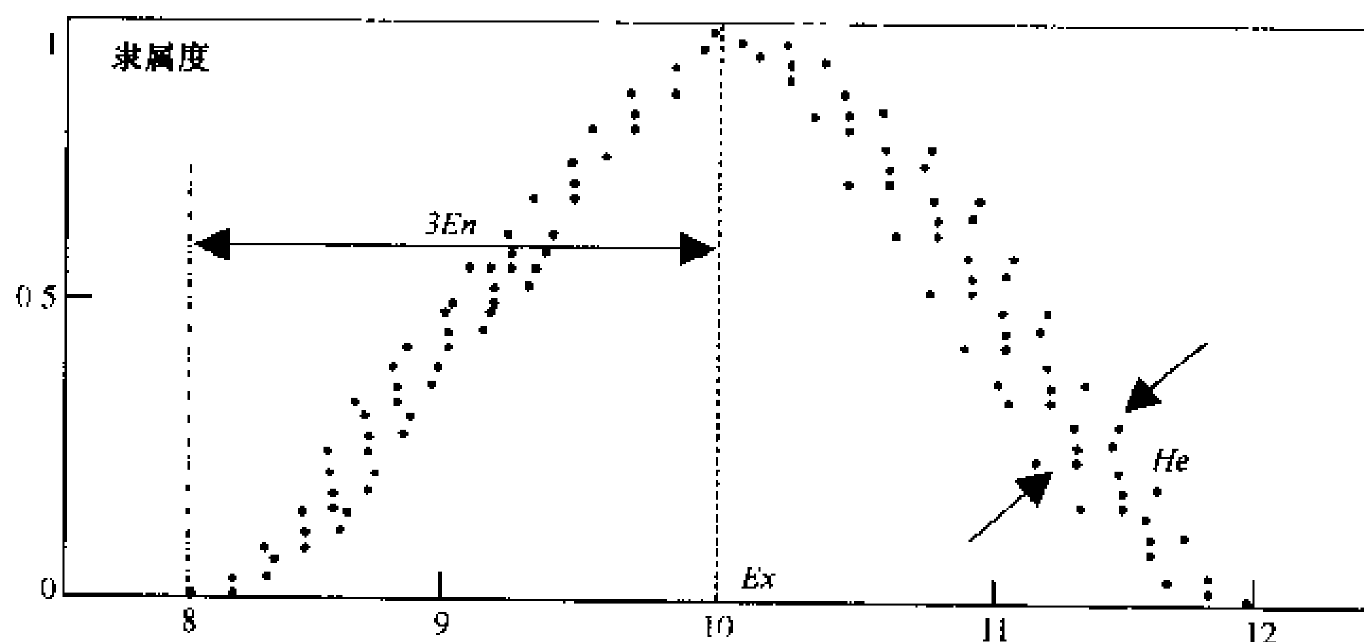


图 23-3　云的数学特征表示示意图

正态云是最重要的云模型。一维正态云的数学期望曲线 MEC 为：

$$\mathrm{MEC}_A(x) = \exp\left[-\frac{(x-Ex)^2}{2En^2}\right] \tag{23-2}$$

一维正态云的生成过程为：

（1）生成以 Ex 为期望值、En 为标准差的正态随机数 x_i：

$$x_i = G(Ex, En) \tag{23-3}$$

（2）生成以 En 为期望值、He 为标准差的正态随机数 En'_i：

$$En'_i = G(En, He) \tag{23-4}$$

（3）令（x_i，u_i）为云滴，计算 u_i：

$$u_i = \exp\left[-\frac{(x_i - Ex)^2}{2En'^2_i}\right] \tag{23-5}$$

二维和多维正态云模型，以及 Γ 云、三角云、梯形云等云模型的生成算法详见邸凯昌（2001）。云的生成可以用软件的方式来实现，也可以固化在硬件中，称为云发生器（cloud generator）。

邸凯昌把基于云模型的属性泛化方法与关联规则法的 Apriori 算法结合起来，实现了从空间数据库中发现关联规则（Di K C. et al.，1998；邸凯昌，2001）。整个过程分两步：首先作属性泛化，然后在泛化后的数据上实施 Apriori 算法，也就是把基于云模

型的属性泛化方法作为Apriori算法的预处理手段。云模型与Apriori算法结合方法及其在经济地理中的应用详见邸凯昌（2001）。

方兆宝等（2003）围绕卫星红外遥感影像提取青藏高原上夏季中尺度对流系统（MCSs）的特征，研究了流形群运动目标自动识别与跟踪匹配技术。

23.2.3 人工智能类算法

1. 诱导学习法（Inductive Learning）

诱导学习法是一种重要的DMKD方法，旨在从大量的经验数据中归纳抽取一般的判定规则和模式。诱导学习法来源于机器学习领域，是该领域最核心、最成熟的分支。诱导学习法的一般模式是：

（1）已知：① 关于某类目标中个别具体目标的状况、过程等的描述（事实）F；② 关于目标泛化描述的初始归纳断言 H_0（可空）；③ 有关背景知识 B，包括关于领域知识、约束条件和假设等。

（2）求断言 H，H 应蕴含事实 F 并满足背景知识 B：

$$F \mid< H \quad (F\text{ 泛化为 }H\text{, 非保真归纳推理})$$

$$H \mid> F \quad (H\text{ 具体化为 }F\text{, 保真演绎推理})$$

诱导学习法根据有无指导或监督而分为监督学习（supervised learning）和非监督学习（unsupervised learning）。所谓监督学习，是有训练例子（经验数据）的分类；而非监督学习则是无训练例子的分类。目前，诱导学习法的算法有很多种，其中最著名的算法是ID3（Quinlan J. R.，1983，1986）和C4.5（Quinlan J. R.，1993）。C4.5算法是一种决策树方法，由ID3算法发展而来，在ID3的基础上增加了将决策树转化为等价产生规则的功能，并解决了连续取值的数据学习问题。两者均采用熵来选择属性，分类速度快，适合大数据库的学习和遥感图像分类。

此外，Han J. 等（1992）提出了一种面向属性的诱导学习法（attribute oriented induction，AOI），专门用于从数据库中发现知识。该法通过扩展的SQL来表达学习任务，包括与任务有关的数据、背景知识、所需阈值、知识表达方式等；并通过查询、投影等关系操作，将与任务相关的数据提取出来，形成初始的关系表，并在表中增加一个计数属性，其初始值为1。诱导学习首先在初始关系表中进行，将属性值用概念树中的高一层概念来代替，并合并相同元组，刷新计数属性，得到泛化的关系表。若某项属性没有高层概念，则删除。重复进行这一过程，直到泛化表中元组的个数以及属性值的个数小于相应的阈值。最后，将泛化关系表转化为规则。

AOI法所需的背景知识以概念树的形式给出；通过概念树的提升对数据进行概括和综合，进而归纳出高层次的模式或特征。该法所获取的规则与前述算法得到的规则相比更加概括和抽象，适合人们阅读和理解，因而更适合SDMKD和决策支持。GIS数据库中，有属性概念树和空间关系概念树两类。背景知识可以由用户提供，也可以作为知识发现任务的一部分而自动获取。

基本的AOI法用于特征规则（characteristic rules）提取。其输入为关系数据库、概念树和学习任务（目标类、知识表达方式等）；输出为特征规则。算法过程为（Han J.

et al., 1992, 1993)：

(1) 选择与任务相关的数据，形成初始关系表；

(2) 实施基本的 AOI 算法，得到最终的泛化关系表；

(3) 将最终的泛化关系表转化为特征规则。

其中第 (2) 步的详细过程如下：

```
Begin
  For each attribute Ai (1 = 1 ~ n, n 为属性个数) do
    While Ai 中属性值个数大于阈值 do
      {
      If 概念树中没有 Ai 的高层概念
      Then 删除 Ai
      Else 用 Ai 的高层概念代替 Ai 的属性值，合并相同元组，刷新计数属性
    }
    While 泛化关系表中元组个数大于阈值 do
      {
      选择需进一步泛化的属性继续泛化，合并相同元组，刷新计数属性
    }
  End
```

对基本的 AOI 算法稍作修改，就可以用于提取除特征规则以外的其他多种类型的规则。例如，若要提取区分规则 (discrimination rules)，只需在学习任务中指明目标类和对比类，对两类同时实施 AOI 算法，并在泛化关系表中去掉目标类和对比属性相同的元组即可。AOI 算法有效、高效的关键是提供好的概念层次，这样才能充分地利用背景知识直接面向属性进行诱导学习。

2. 神经网络法 (neural network approach)

神经网络是由多个简单的处理单元（神经元）按某种方式相互连接而成，动态响应外部的输入信息。神经网络具有自组织、自适应学习功能，降低了传统模式识别方法对约束条件的需要，因而对某些知识识别问题具有很大的优越性，主要用于获取分类知识。但是，由于神经网络法在获取分类知识时要多次扫描训练数据，网络训练时间比较长；而且所获取的知识是隐含在网络结构中的，并不是显示地表达为规则，因而不易被人们理解和解释。

鉴此，Lu H. J. 等 (1995) 提出了一种用神经网络获取分类规则的算法，称为 NeuroRule。该算法采用三层 BP 网络，由网络训练、网络剪枝和规则提取 3 步构成。其中网络训练与常规的 BP 网络训练基本相同；网络剪枝的目的是在不增加错分率的前提下，去掉冗余的结点和连结，得到一个简练的网络；然后在此网络中，根据输入值、隐含结点值、隐含结点激活值及输出值之间的依赖关系来提取规则。典型试验表明，在分类精度相同的情况下，NeuroRule 法比 C4.5 法提取的规则更为精练。

神经网络法用于遥感影像分类已经比较成熟，在 GIS 数据分类、聚类、预测等知识获取方面的研究与应用也已开展。相信，只要充分利用神经网络的优点，结合或借鉴

其他方法来克服其缺点，必然会使神经网络法在 SDMKD 中得到广泛应用。

3. 遗传基因算法（genetic algorithms，GA）

遗传基因算法最早是由 John H.（1975）提出，是一种解决最优化问题的有效方法。该法的基本思想是：仿效生物的进化与遗传，根据“生存竞争”和“优胜劣汰”的原则，通过复制、教化和突变等操作，使问题的初始解一步一步地逼近最优解。由于遗传基因算法具有智能搜索、渐进优化、易获全局最优解、黑箱式结构等特点，而且适于并行计算，因而近年在机器学习、图像处理、模式识别、自动控制和社会科学等领域得到越来越广泛的运用。

DMKD 和 SDMKD 中的许多问题，如分类、聚类、预测等知识获取问题，均可以表达或转换为最优化问题，进而可以采用遗传算法来解决。目前，这方面的研究还不够成熟，像知识表达、定性定量转换、不确定性推理等问题还有待研究解决。

4. 粗集法（rough sets approach）

粗集理论是波兰华沙大学 Pawlak Z. 教授于 1982 年提出的一种智能决策分析工具（Pawlak Z.，1982），已被广泛用于不精确、不确定、不完全的信息分类分析和知识获取。与其他知识发现方法相比，粗集法可以在 GIS 数据库中数据不确定的情况下获取多种知识，其应用方向包括：GIS 数据属性表的一致性分析、属性重要性、属性依赖、属性表简化、最小决策和分类算法生成等，粗集理论与其他知识发现算法相结合，可以在空间数据库中数据不确定的情况下获取多种知识（邸凯昌，2001；周海燕等，2002）。

设 U 是非空论域；R 为一个等价关系（equivalence relation）；$A=(U,R)$ 称为一个近似空间（approximation space）；U/R 表示 R 中所有等价类的集合或 U 的分类；$[x]_R$ 表示 R 中包含 x 的等价类（equivalence class）；R 中的等价类称为基本集（elementary set）；基本集的有限并集称为可定义集（definable set）或合成集（composed set）。

设 X 是 U 的子集，$X \subseteq U$，则 X 可用可定义集的术语从 A 中定义粗集的两个重要概念，即 X 的下近似和 X 的上近似：

A 中包含在 X 中的最大可定义集称为 X 的下近似（lower approximation）：

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\} \tag{23-6}$$

A 中包含 X 的最小可定义集称为 X 的上近似（upper approximation）：

$$\bar{R}X = \{x \in U \mid [x]_R \cap X \neq \Phi\} \tag{23-7}$$

进一步定义 BND（X）$= \bar{R}X - \underline{R}X$ 为 X 的边界或边界区；POS（X）$= \underline{R}X$ 为 X 的正区，NEG（X）$= U - \bar{R}X$ 为 X 的负区。

若把 R 看成分类知识，则 X 的正区中的目标可以确定地分类为 X 的成员；负区中的目标可以确定地判断不属于 X，即属于 X 的补集 $-X$；边界区中的目标则无法确定地判断到底属于 X 还是-X。当且仅当 $\overline{R}X = \underline{R}X$，$X$ 是 R 可定义的；当且仅当 $\overline{R}X \neq \underline{R}X$，$X$ 对于 R 是粗集。

于是，得到一个用下近似和上近似定义的子集，该子集称为粗集。可以用图 23-4 形象地说明粗集的基本概念。图中，曲线所包含的部分为集合 X，由矩形单元构成二

维近似空间，其中每个单元均为基本集。

粗集理论认为知识是以对目标的分类能力为基础的。目标可以是指真实的物体、状态或抽象的概念、过程、时刻等。知识包括两部分内容：一是感兴趣领域中的各种分类模式构成的族（Family），二是由明显的知识派生出不明显的事实的推理能力。设 U 是感兴趣的对象非空论域，任意子集 $X \subseteq U$ 可以看成为 U 的一个概念，空集也是一个概念；U 中任何概念族构成对 U 的一个特定分类，即是关于 U 的知识；R 是 U 上的等价关系族，则 $K=(U, R)$ 称为知识库。进而将上述关于集合近似的定义推广至关于分类近似的定义：

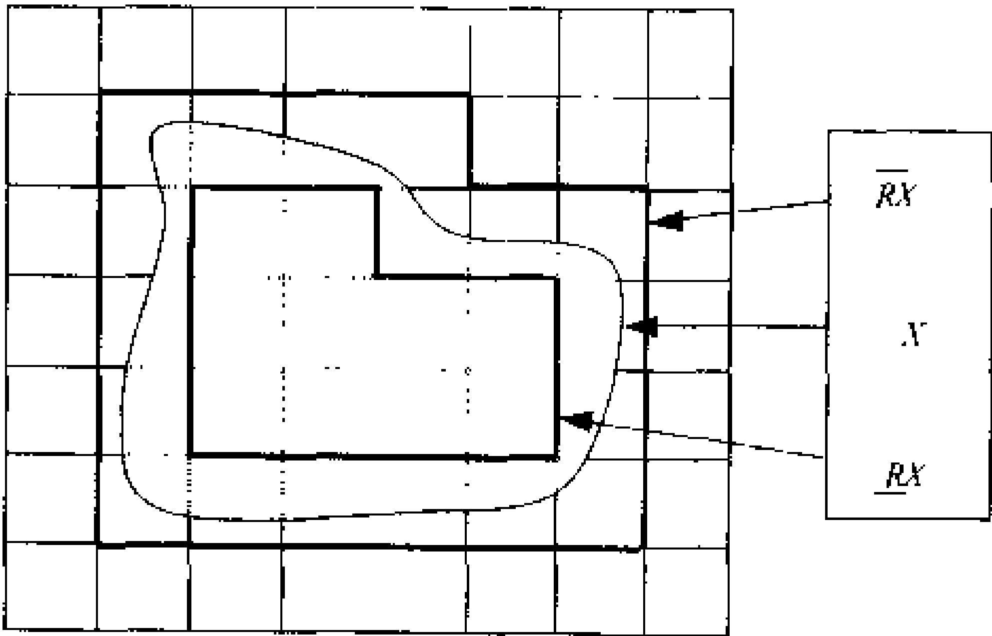


图 23-4　粗集的基本概念
（据邸凯昌，2001）

设 $F=\{X_1, X_2, \cdots, X_n\}$ 是一个非空分类集合族，则

$$\underline{R}F=\{\underline{R}X_1, \underline{R}X_2, \cdots, \underline{R}X_n\} \tag{23-8}$$

$$\overline{R}F=\{\overline{R}X_1, \overline{R}X_2, \cdots, RX_n\} \tag{23-9}$$

分别称为族 F 的 R-下近似和 R-上近似。

并定义两个测度来描述近似分类的不精确性，即 F 的近似分类精度 $\alpha_R(F)$ 和近似分类质量 $\gamma_R(F)$：

$$\alpha_R(F)=\frac{\sum \mathrm{card}\, \underline{R}X_i}{\sum \mathrm{card}\, \overline{R}X_i} \tag{23-10}$$

$$\gamma_R(F)=\frac{\sum \mathrm{card}\, \underline{R}X_i}{\sum \mathrm{card}\, U} \tag{23-11}$$

近似分类精度表示当使用知识 R 对目标进行分类时，可能的决策中正确决策的百分比；近似分类质量表示当使用知识 R 能确切地划入 F 类的对象的百分比。

邸凯昌等（2001）将粗集算法引入 GIS 领域，归纳整理出了粗集算法用于 GIS 中属性分析和知识发现的一整套方法，包括：GIS 数据库中属性表的一致性分析、属性的重要性、属性依赖、属性表简化、最小决策和分类算法生成等，为 SDMKD 开辟了一条新的、有意义的途径。

5. 粗集法与云理论的结合（rough sets approach plus cloud theory）

由于粗集理论强调的是知识的不完全性和不可分辨性，而处理的方法是确定性的，即要求属性值都是定性的。实际上，大量存在的是定量数据，要设法将定量数据转换为定性数据，才能运用粗集方法。云理论恰好提供了定性定量之间转换的模型与工具，因此，云理论可以与粗集算法结合起来进行 SDMKD。云理论与粗集算法结合有两种模式：

（1）先将云理论从定量到定性的转换作为粗集的预处理手段；然后应用粗集算法发现分类决策知识；最后再用基于云理论的不确定性推理方法应用新获得的知识推理出新的定性定量结果。其整个过程如图 23-5 所示。

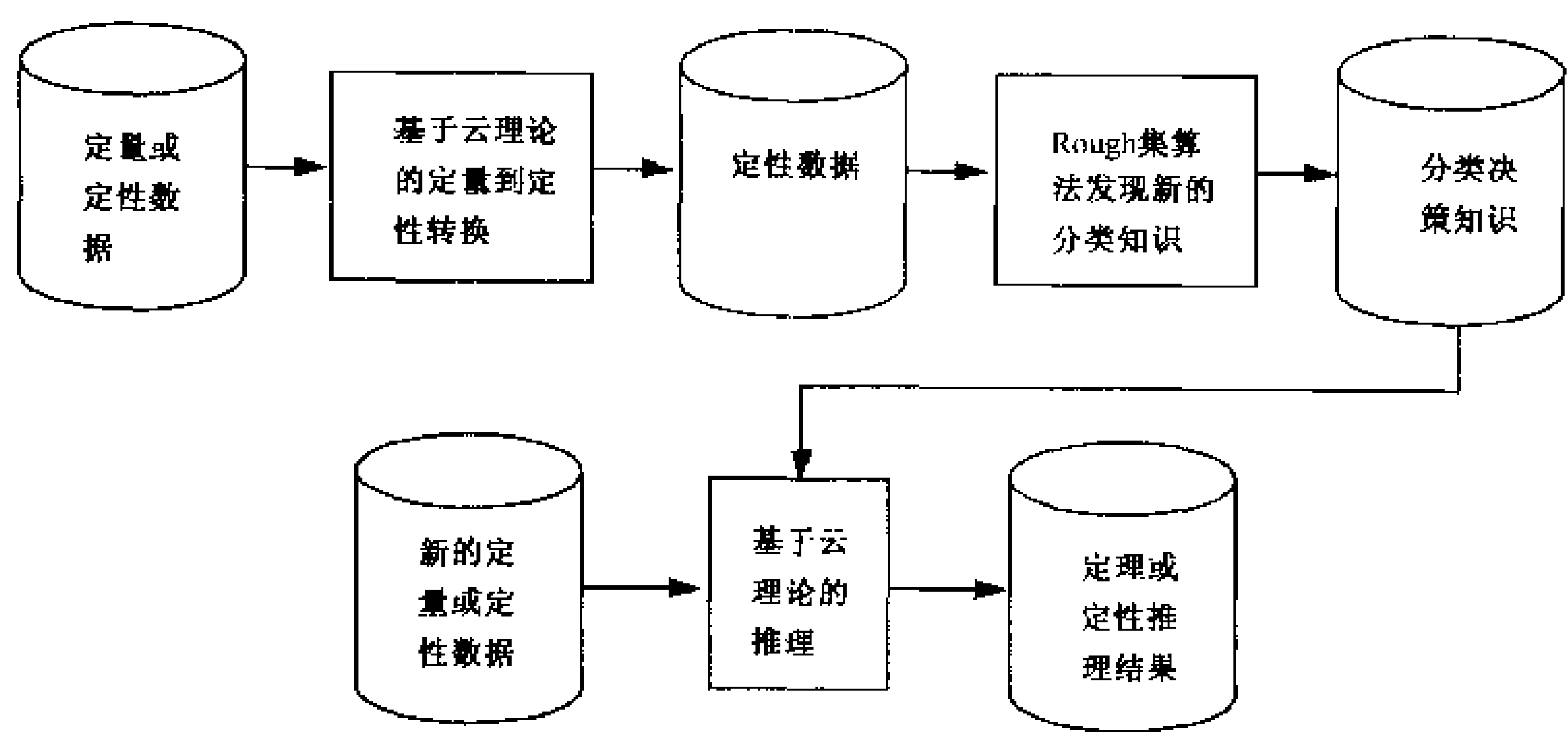


图 23-5　云理论与粗集算法结合的一种典型模式
（据邸凯昌，2001）

（2）先用云理论的一系列方法发现定性知识，然后用粗集算法对所发现的知识进一步简化和精练。

23.2.4　图形图像类算法

1. 图像分析与模式识别（image analysis and pattern recognition）

空间数据库中含有大量的图形图像数据，一些行之有效的图像分析和模式识别算法可以直接用于知识发现，或作为其他知识发现方法的预处理手段。尤其在多源遥感数据融合过程中，如何将同一地区的多传感器、多时相、多光谱影像数据中的有用信息有效地聚合起来，使后续的遥感影像制图或解译工作能充分利用这些信息，图像分析和模式识别算法大有用武之地。可用于图像分析与模式识别的方法主要有：决策树（decision tree）方法、神经元网络（artificial neural network）方法、数学形态学方法和图论方法等（周海燕等，2002）。

为了在空间数据库中发现知识，尤其是发现普遍的几何知识，需要把空间目标的几

何特征纳入到数据挖掘算法 空间目标的基本几何特征包括：线状目标的长度、宽度，面状目标的面积、周长、几何中心等。空间目标的其他形态特征需要用专门的图形图像算法计算，此时，图像分析与模式识别算法就成为 SDMKD 的不可缺少的预处理手段。空间目标的其他形态特征包括：

(1) 点群的形态特征（Zhang T. et al.，1997)：点群的重心 X_0（点群中的点的平均坐标)、半径 R（点群中的点到重心的平均距离)、直径 D（点群中的点对间的平均距离)；

(2) 线状目标的形态特征：线状目标的复杂度（或称曲折度，可用分数维表示)、总体延伸方向等；

(3) 面状目标的形态特征：面状目标的密集度（$\frac{P}{4\pi A}$，A 为面积，P 为周长；圆形为 1)、体态比（最小外接矩形的宽长比；圆形和正方形为 1)、凸度（图形面积与其凸包面积之比；凸图形为 1，凹图形小于 1)、主轴方向（面状目标的总体延伸方向)、占有率（图形面积与其最小外接矩形面积之比；矩形为 1)、图形的直径（图形内任意两个点对之间距离的最大值）等。

2．可视化法（Visualization)

可视化是一种将数据（特别是多维数据）以图形方式显示的计算机技术，其发展分支包括科学可视化（scientific visualization）和虚拟现实（virtual reality)。它拓宽了传统的图表功能，使用户对数据的剖析更清楚。由于人类对于图形的模式识别能力非常强大，很容易从各种图形图像表示中直接发现规律或异常，远远超过现有的任何模式识别与异常检测的计算机技术。针对这一特点，计算机所要做的只是把大量的多维信息以多种方式可视化出来，剩下的就交由人的思维来完成最终的知识发现。

事实上，GIS 本身就是一个可视化系统，尤其随着 2D GIS 向 2.5D、2.75D、3D 和 4D 方向发展，GIS 对多维动态时空信息的表达与可视化能力不断增强，必然会给 SDMKD 带来更多的便利和途径。

以上所述 DMKD 与 SDMKD 算法并不完全，有些算法还可以进一步综合起来使用。比如，邸凯昌（2001）将 EDA/ESA、AOI 和粗集等结合起来，提出了一种探测性诱导学习方法（exploratory inductive learning，EIL)。该方法的核心是 AOI，其作用是从空间数据库中得到泛化的数据；EDA/ESA 用于交互地提取与任务相关的数据；粗集算法则用于泛化数据的进一步简化和决策算法生成，使得在保持普遍化数据内涵的前提下最大限度地精练知识。有兴趣的读者可以查阅有关资料。

23.2.5 空间分析类算法

空间分析是 GIS 的关键技术，也是 GIS 区别于其他图形系统的主要标志之一。目前常用的 GIS 空间分析功能有：拓扑结构分析、空间缓冲分析、距离分析、密度分析、叠置分析、地形分析、网络分析等。应用这些分析算法可以交互式地发现目标在空间上的相连、相邻、共生等关联关系及一些空间统计特征。空间分析技术往往是应用领域知

识来产生新的空间数据，而数据挖掘是要从大量的空间数据中发现知识。因此，空间分析技术常作为预处理和特征提取方法与其他数据挖掘算法结合起来，实现从空间数据库中发现知识，并提高自动化和智能化水平。

以Ester提出的基于邻域图（neighborhood graphs）和邻域路径（neighborhood paths）的挖掘算法为例，Ester将一个空间特征定义为空间数据库中具有空间与非空间属性性质的目标对象集，并以非空间属性值出现的相对频率和不同空间对象出现大的相对频率（目标对象集相对于整个空间数据库）作为感兴趣的性质，从空间目标集合经过它的邻域扩展后的集合中，发现相对频率的明显不同，以此提取空间规则（据周海燕等，2002）。邸凯昌等（1999）则将所提出的EDA技术与空间分析结合起来，构成探测性空间分析技术（exploratory spatial analysis，ESA），使得空间分析更加灵活方便。

总之，SDM的发展十分迅速，已经取得了可喜的成果，但仍有大量的理论与技术问题有待于进一步研究和探索，如多分辨率的空间数据挖掘、并行空间数据挖掘算法、多媒体空间数据挖掘技术、基于空间元数据的SDM、知识的可视化表达与人机交互、分布式空间数据库的知识发现、空间数据挖掘语言、SDM与GIS、ES的集成、SDM与其他IT新技术（如不确定性、OLAP、模糊技术等）的集成等。

§23.3 矿山空间数据挖掘系统设计

矿山数据挖掘（mine data mining，MDM）就是从海量的矿山数据中挖掘、发现矿山系统中内在的、有价值的信息、规律和知识的过程。这些信息、规律和知识可对矿山的安全、生产、经营与管理发挥预测和指导作用。作为数字矿山有机组成部分的MDM技术与三维地学模拟（3D GM）一起并称为数字矿山的包装系统（吴立新，2002），没有先进的矿山数据挖掘技术就无法展示矿山多源数据中蕴含的内在信息和规律，也就称不上真正的数字矿山。

23.3.1 矿山数据的基本特征

矿山GIS中的时空数据具有以下基本特征：

（1）多源异质：矿山数据的获取手段与方法多种多样，包括：地面全站仪测量、井下测量、航测、GPS、遥感、数字摄影测量、勘探（钻探、槽探、山地工程、化探、物探）、监控、传感、图纸数字化、图式的扫描矢量化等等。数据的来源不同，属性、结构类型、格式、精度、编码等都存在较大差异，且分别存储在不同的平台和不同形式的数据文件、数据库中。

（2）强时空变化：由于矿业生产的动态性，矿山数据表现出很强的时空变化特征。绝大部分矿山数据具有3D空间维和时间维。例如：煤层上覆岩层、地表、地面建筑物等随着工作面的推进，而发生弯曲、断裂、沉陷、破坏，煤层和各种巷道峒室也随工作面和巷道的推进而发生显式变化。

（3）海量性：前面已提到，矿山数据采集方法是多种多样的，存储、分布在不同的平台和各种形式的数据文件或数据库中，这就使得矿山数据规模非常庞大。

(4) 多尺度性：通过遥感、摄影测量获取的整个煤田、矿区的数据；通过地面测量测得的矿山数据，以及井下测量得到的数据；通过勘探获取的数据，这些数据在空间尺度（粗度）上有巨大的差距。

由于矿山数据的以上特性，使得要对它进行有效的组织、管理非常困难，是一般数据库无法胜任的。因此，矿山数据挖掘必须以空间数据仓库技术为依托。

23.3.2 矿山数据仓库体系结构

矿山空间数据仓库管理着海量的矿山地上及地下地物、地貌的几何信息、拓扑信息和属性信息，以多维方式来组织数据，负责数据的简化、压缩和存储，同时负责数据的检索、提取和维护功能。空间维和时间维是矿山空间数据仓库反映矿山动态变化的基础，它们的数据组织方式是整个空间数据仓库技术的关键。矿山空间数据仓库的数据组织方式可以是基于关系表的存储方式或多维数据库存储方式。基于关系表的数据模型主要有星型和雪花模型；多维数据库数据模型主要是超立方体结构模型。

通常，数字矿山中的空间数据仓库是由集团公司或矿务局建立的。为了高效地对整个矿山不同地点、不同来源、不同格式的数据进行组织、维护、管理，同时满足不同级别请求、决策的需要，宜采用“反馈式并行开发”模式，如图 23-6 所示。各矿空间数据库中的数据是直接从现场采集得到，业务处室对来自各矿的部分空间数据库数据进行分类、加工，在此两者的基础上提取到全局空间数据仓库中。全局空间数据仓库是按主题构建的，其目的是为空间决策支持提供信息基础，为整个矿区的高层决策服务。

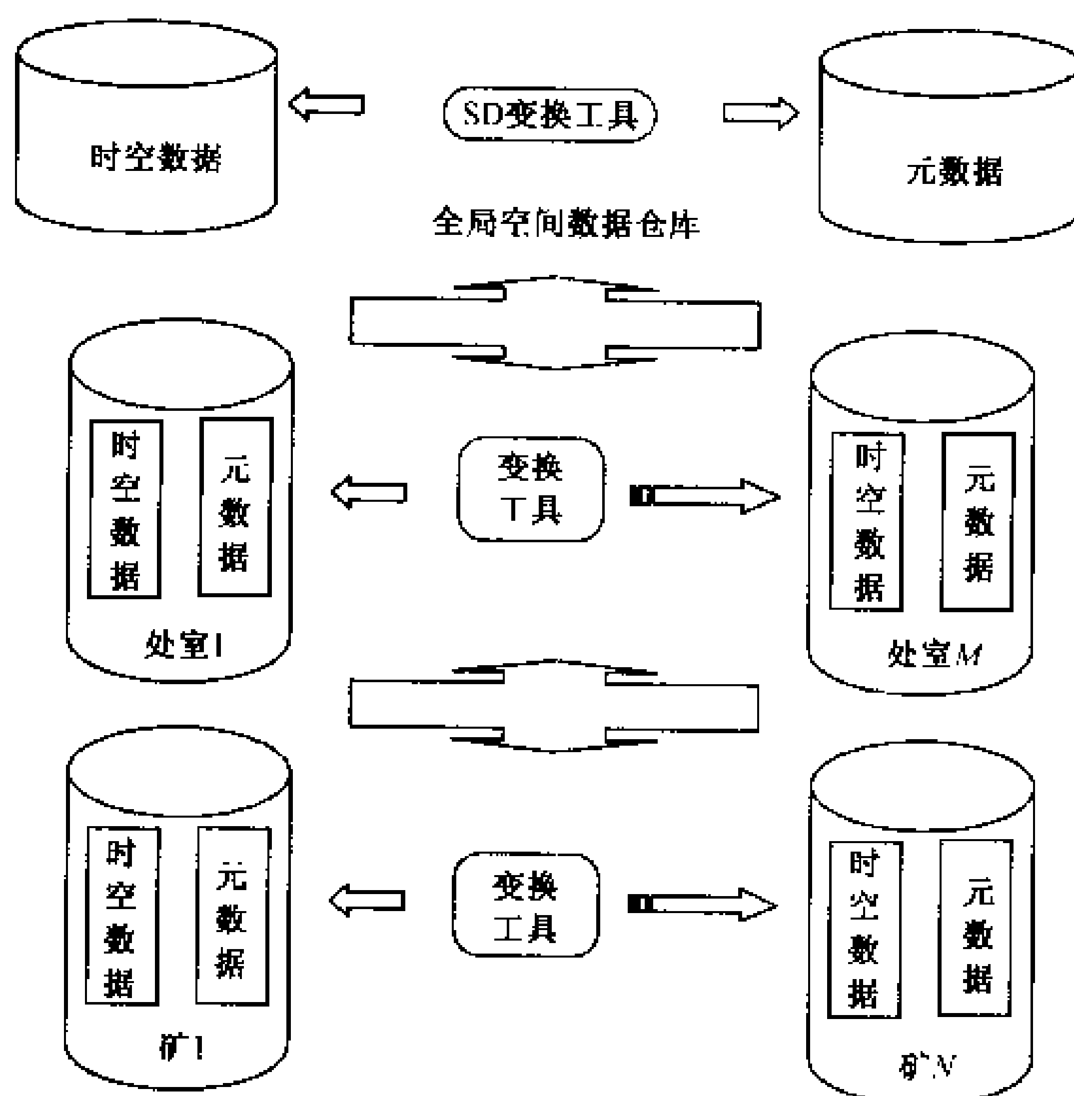


图 23-6 矿山空间数据仓库体系结构

(1) 时空数据：这是空间数据仓库的主体，如前所述，它们来自不同获取手段，并且分别存储、分布在微机、工作站、网络服务器等不同的平台及各种形式的数据文件和数据库中。

(2) 元数据：定义了数据的存贮模型和数据结构以及转换规则和控制信息等，是形成空间数据仓库的关键所在，离开了元数据，空间数据仓库中的数据就无法形成空间数据仓库，而成为垃圾数据。元数据包括三部分：元数据标准、元数据操作工具和元数据库。依据元数据从各种数据源（库）中提取数据形成空间数据仓库的工作模式如图23-7所示。根据矿山数据挖掘对元数据的要求及分布式空间数据库中数据的特点制定出矿山元数据的标准，依据此标准建立元数据操作工具及元数据库，三者共同组成矿山元数据。以此元数据为依托，按照一定的主题要求，对各分布式数据库中的数据进行选择、分析、提取、存储，形成空间数据仓库，为矿山数据挖掘提供数据源。

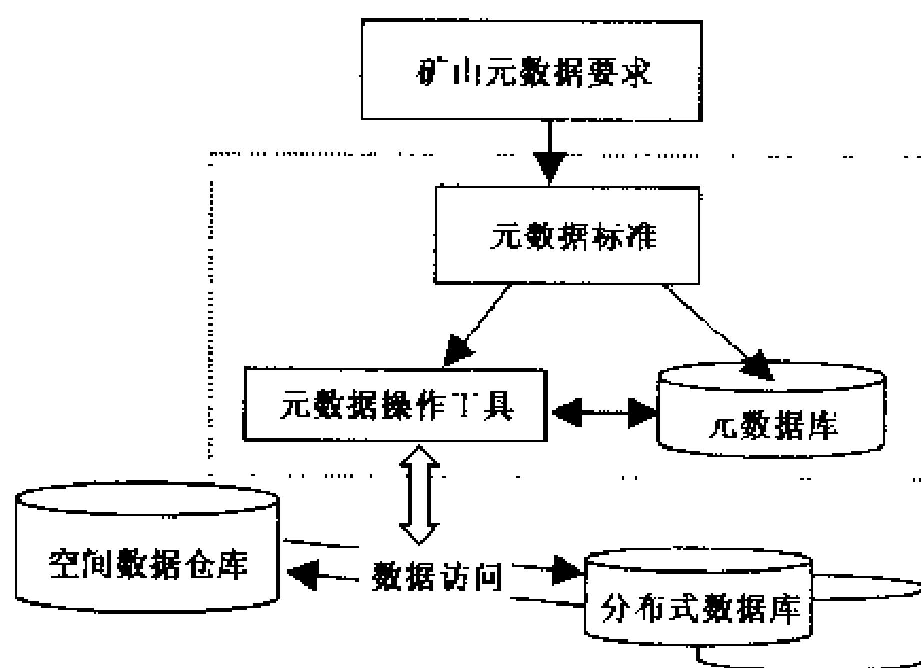


图 23-7 数据挖掘中元数据的工作模式

(3) 数据变换工具：为了优化空间数据仓库的分析性能，时空数据必须经过加工、变换后方可进入数据仓库；变换包括提炼、转换、空间变换。其中数据转换工具是各数据源通向矿山数据仓库的桥梁，使源数据在格式上取得了统一。

23.3.3 矿山数据挖掘过程

鉴于矿山数据的复杂性，参照数据挖掘技术的一般过程，基于空间数据仓库的矿山数据挖掘可分为 5 个阶段，如图 23-8 所示。

1. 矿山数据采集和发送

将各数据采集系统中采集到的数据通过一定的方式发送、存储到各个分布式数据源（库）中。利用数字化仪、扫描仪，将矿山图纸输入计算机；遥感、数字摄影测量的结果数据通过磁带导入计算机；对于从外业测绘仪器如全站仪、GPS 接收机、光电测距仪获取的野外和井下测量数据，可利用数据传输接口通过相应的数据识别系统输入微机；物探数据可由检波器直接数字化，存储在单元存储器中，然后记录设备指令，把数

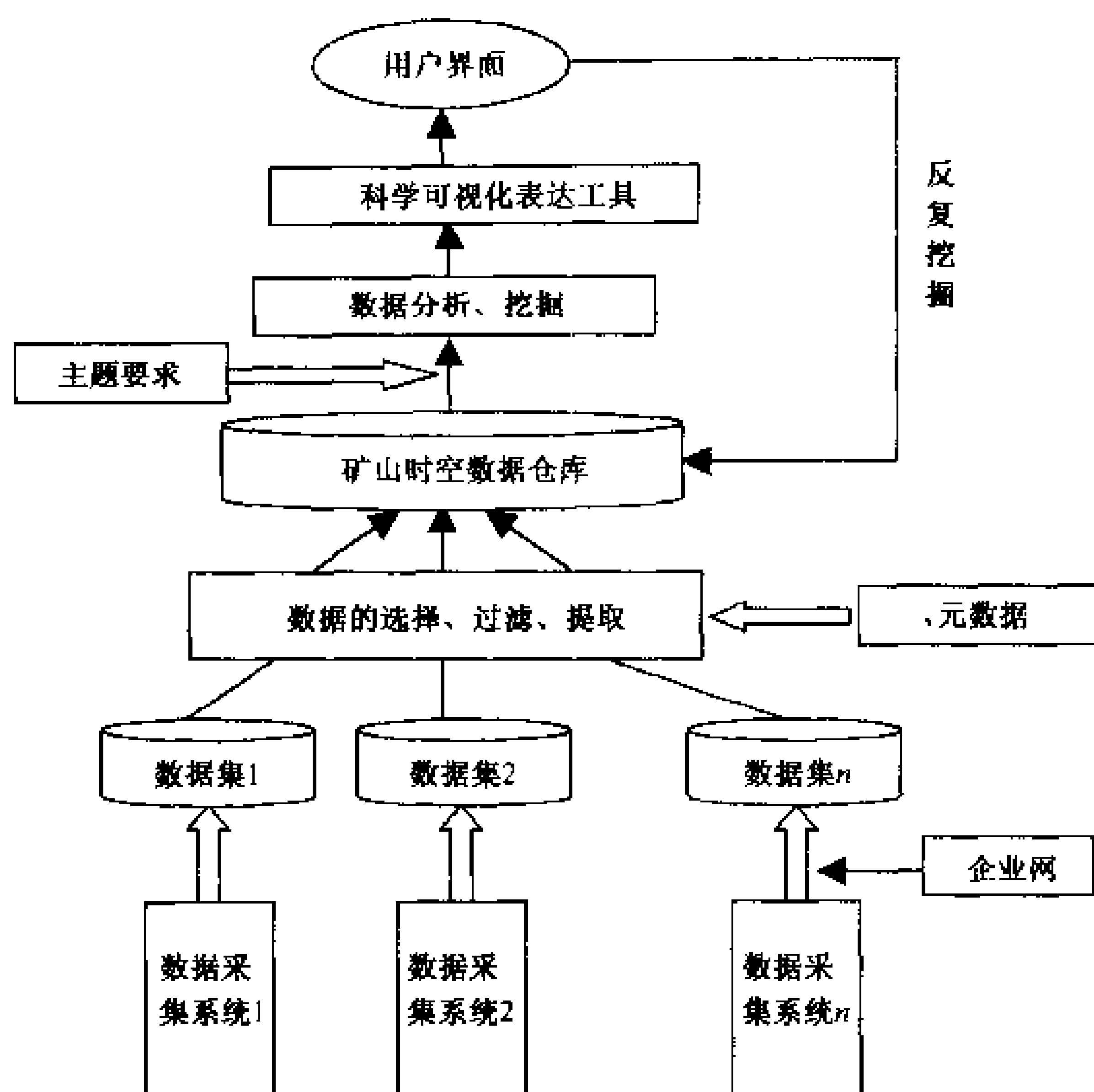


图 23-8 矿山空间数据挖掘过程

据传送到车载计算机中；钻探和槽探数据首先由人工记录，再利用专用软件，例如：用GeoFrame和GEMA等成图并存入计算机中。本地计算机经过企业网，把这些数据发送到数据源服务器中，存入相应数据库或数据文件中。

2．矿山数据的主题构建

根据实际需求，确定主题域，包括两方面内容：① 确定主题，是安全生产、通风管理、产品销售还是地面建筑物保护等等；② 确定相关人员，例如：决策者、管理者、生产者、技术人员等。随后，按照主题要求为每个主题抽取数据，在系统的帮助下建模，从现有的分布式空间数据库中提取数据，并对其进行选择、过滤、提取形成空间数据仓库。

3．从空间数据仓库中挖掘信息

根据确定的主题和任务，选择合适的数据模型和适当的数据挖掘算法，从空间数据仓库中分析、提取、挖掘出所需的知识信息。数据挖掘的方法很多，常用的有：数据总结、统计分析、空间关联、序列模式分析、分类分析和聚类分析。

此外，近年来又涌现出了一些新方法，例如：云模型法、模糊模式判别法以及遗传算法等等。在实际数据挖掘中非常重要的一点是选择适当的数据挖掘算法，主要取决于问题的类型以及数据的类型和规模。而利用这些技术相互配合可以取得更好的效果。

4. 科学可视化表达

数据挖掘的最后一个重要的环节是运用科学可视化技术，将挖掘所得到的知识以易于理解的方式直观地表达出来，最大限度地利用信息，实现信息共享。

矿山数据挖掘可视化的方向是以 3D GM 和虚拟矿山技术为背景，叠加上从数据仓库中挖掘出的信息的可视化影像，生动、形象、逼真地呈现在决策者和管理者面前，为矿山的资源评估、规划设计、开拓设计、安全生产和管理决策服务。

用 3D GM 和虚拟矿山技术将整个矿山全貌展现出来，包括地面状况、井下各水平主要巷道、硐室，各采掘工作面以及地层及其构造等等，再可选择地加入某些挖掘出的信息影像。管理者和工程技术人员就可以在浸入式的虚拟环境中交互、体验、身临其境，感悟出挖掘到的信息，提高矿山安全生产与决策服务的信息化水平。

5. 对挖掘的知识进行评价

首先对其进行一致性检查，以确定本次挖掘的知识与以前挖掘的知识是否相互抵触。其次，判断受挖掘的知识是否满足要求，以决定是否需要反复挖掘以获取更加有价值的知识、信息。在这个阶段中，不同的应用领域、不同的学科有各自的评估标准和评估方法。对于矿山而言，评估标准和方法基本还是空白，有大量的工作急需去做。

总之，矿山数据挖掘技术是一个极有挑战性的研究领域，其技术关键主要是空间数据仓库的实现以及矿山空间数据仓库元数据标准的建立。

参 考 文 献

邸凯昌，李德仁．1996．KDD 技术及其在 GIS 中的应用与扩展．中国 GIS 协会第二届年会，北京，1996 年 9 月

邸凯昌，李德仁，李德毅．1997．空间数据挖掘和知识发现的框架．武汉测绘科技大学学报，22（4）：328－332

邸凯昌，李德仁，李德毅．1999．云理论及其在空间数据发掘和知识发现中的应用．中国图像图形学报，4（11）：930－935

邸凯昌，李德仁，李德毅．1999．用探测性的归纳学习方法从空间数据中发现知识．中国图像图形学报，4（11）：924－929

邸凯昌．2001．空间数据发掘与知识发现．武汉大学出版社．42－165

杜明义，古寿松，郭达志．2000．基于空间数据仓库的数据采掘．计算机工程与应用．（11）：32－34

常占强，吴立新，漆小龙，陈书琳．2002，基于空间数据仓库的矿山数据挖掘技术探讨．第六届全国矿山测量学术会议论文集，45～48

郭宜斌．1996，数据仓库技术的基本概念和发展现状．微电脑世界，（4）

胡 侃，夏绍玮．1998．基于大型数据仓库的数据采掘：研究综述．软件学报．9（1）：53－63

李德仁，程 涛．1995．从 GIS 数据库中发现知识．测绘学报．22（4）：37－43

李德毅，史雪梅，孟海军．1995．隶属云和隶属云发生器．计算机研究与发展，42（8）：32－41

李永平，陈意云，黄刘生．1998．数据采掘技术回顾．小型微型计算机系统．19（4）：74－81

刘东波．2000．数据仓库技术的现状与未来．微型机与应用．（7）：4－7

史金红，吴永明．2000．数据仓库中元数据的管理．电子工程师．（2）：9－12

苏 辉，吴立新，陆镇虹，王金庄．2002，GIS 技术在电信领域中的应用及应解决的问题．电信科学，8（2）：28－31

苏 辉，吴立新，陆镇虹，王金庄．2002，基于 GIS 和 ES 的光纤接入网规划系统的设计和实现．地理学与国土研究 18（1）：19－22

汪 闽，周成虎．2002．空间数据挖掘方法的研究进展．地理信息世界．（2）：26－31

王　珊，罗　立．1996，从数据库到数据仓库．计算机世界，28（专题版）
王晋年等　1999．以地物识别和分类为目标的高光谱数据挖掘．中国图像图形学报，4（11）：957－964
吴立新等．2000．论21世纪的矿山--数字矿山．煤炭学报．25（4）：337－342
吴立新等．2002．三维地学模拟与虚拟矿山系统．测绘学报．31（1）：28－33
杨超伟，李琦等．数字地球中的空间数据仓库研究．中国图像图形学报，1999（增刊）
周海燕，王家耀，吴　升．2002．空间数据挖掘技术及其应用．地理信息世界．（2）：32～33
赵霈生，杨崇俊．2000．空间数据仓库的技术与实践．遥感学报．4（2）：157－160
Agrawal R., Imielinski T. and Swami A., 1993. Mining association rules between sets of items in large databases. Proc ACM-SIGMOD Int. Conf. Management of Data, 207～216
Agrawal R. and Srikant R., 1994. Fast algorithms for mining association rules in large databases. Processing of the 20th International Conference on Very Large Data Bases, September, 478－499
Anand S. S., Bell D. A. and Hughes J. G., 1996. EDM: A general framework for data mining based on evidence theory. Data & Knowledge, 18: 189～223
Cheung D. W., Han J., Ng V., and Wong Y. C., 1996a Maintenance of discovered association rules in large databases: An incremental updating technique. In Proc. Int'l Conf. On Data Engineering, New Orleans, Louisiana
Cheung D. W., Ng V. and Tam B. W., 1996b. Maintenance of discovered knowledge: a case in multi-level association rules. In Proc. 2nd Int. Conf. On Knowledge discovery and Data Mining (KDD-96), Protland, Oregon
Cheung D. W., Han J., Ng V., Fu A. W. and Fu Y., 1996c. A fast distributed algorithm for mining association rules In Proc. 4th Int. Conf. on Parallel and Distributed Information System (PDIS-96), Miami Beach
Di K. C., Li D. Y. and Li D. R., 1998a. Knowledge representation and discovery in spatial databases based on cloud theory. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 3/1, Columbus, Ohio, July, 1998
Di K. C., Li D. R. and Li D. Y., 1998b. Intelligent query in spatial databases based on cloud model. In Li D. R. et al., Eds. Spatial Information Science, Technology and Its Applications, WTUSM Press, SIST'98, Wuhan, Dec., 1998
Ester M. et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proc. 2nd Int. Conf. On Knowledge Discovery and Data Mining, Portland, USA
Ester M, Kriegel H. P. and Sander J., 1997. Spatial data mining: a database approach. Proc. 5th Int. Sym Large Spatial Database, Berlin: 47～66
Fayyad U. M., Piatetsky-Shapiro G., Smyth P., and Uthurusamy R. eds., 1996a. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press
Frawley W., Piatetsky-Shapiro G and Matheus C., 1991. Knowledge discovery in databases: an overview In G Piatetsky-Shapiro and W. Frawley (ED), Knowledge Discovery in Databases, AAAI/MIT Press
Han J., Cai Y. and Cercone N., 1992. Knowledge discovery in databases: an attribute-oriented approach. Proc. of 18th VLDB, 547～559
Han J and Cercone N., 1993. Data-driven discovery of quantitative rules in relational databases. IEEE Trans. Knowledge and Data Engineering, 5: 29～40
Han J., Fu Y., 1995a. Discovery of multiple-mevel association rules from large databases (KDD'94), Seattle, WA
Han J., 1995b From database systems to knowledge-based systems: an evolutionary approach Conf. Tutorial at the 11th Int. Conf. On Data Engineering (ICDE'95), Taipei
Han J. and Fu Y., 1995c. Discovery of multiple-level association rules from large databases. Proc of the 21st Int. Conf. on Very Large Data Bases, September, 1995: 420－431
Han J, 1996. Data mining techniques. ACM-SIGMOD'96 Conf Tutorial
Kamber M., Han J. and Chiang J. Y., 1997. Meta rule-guided mining of multi dimensional association rules using data cubes. Proc. of the 3rd Int. Conf. on Knowledge Discovery and Data Mining, Newport Beach, California
Koperski K. and Han J., 1995. Discovery of spatial association rules in geographic information databases. In Proc 4th Int Symp. On Large Spatial Databases (SSD'95), Portland, Maine: 47－66
Koperski K., Adhihary J. and Han J., 1996b. Spatial data mining: progress and challenges. SIGMOD'96 Workshop on

Research Issues on Data Mining and Knowledge Discovery (DMKD'96), Montreal, Canada

Li D. R. and Cheng T., 1994. KDG: knowledge discovery from GIS propositions on the use of KDD in an intelligent GIS. In Proc. ACTES, the Canadian Conf. on GIS

Li D. R., Di K. C. and Li D. Y., 1997. A Framework of spatial data mining and knowledge discovery. In Proc. Int Workshop on Image Analysis and Information Fusion (IAIF'97), Adelaide, Australia

Li D. Y., Han J., Chan E. and Shi X. M., 1997a. Knowledge representation and discovery based on linguistic atoms Proc. of 1st Pacific-Asia Conf. on KDD&DM, Singapore

Li D. Y., Cheung D. W., Shi X. M. and Ng V., 1997b. Uncertainty reasoning based on cloud models in controllers. Computers and Mathematics with Applications, Elsevier Science

Li D. Y., Di K. C., Li D. R. and Shi X. M., 1998. Mining association rules with linguistic cloud models. PAKDD'98, The Second Pacific-Asia Conference on Knowledge Discovery & Data Mining, Melbourne, Australia

Lu H. J., Setiono R. and Liu H., 1995. NeuroRule: a connectionist approach to data mining. Proc. of 21st VLDB Conf., Zurich, Switzerland

Lu W., Han J. and Ooi D. C., 1993. Discovery of General Knowledge in Large Spatial Database. Proc. Far East Workshop on Geographic Information System, Singapore: 275~289

Murray A. T. and Estivill-Castro V., 1998. Cluster discovery techniques for exploratory spatial data analysis. Int. J. GIS, 12 (5): 431~443

Ng R. and Han J., 1994. Efficient and Effective Clustering Method for Spatial Data Mining. Proc. Int. Sym. Very Large Data Bases, Santiago: 144~155

Park J. S., Chen M. S. and Yu P. S., 1995a. An effective hash-based algorithm for mining association rules. In Proc. 1995 ACM-SIGMOD Int. Conf. Management of Data, San Jose

Park J. S., Chen M. S. and Yu P. S., 1995b. Efficient parallel data mining for association rules. Proc. of 4th Int. Conf. on Information and Knowledge Management: 31~36

Pawlak Z., 1982. Rough Sets. Int. J. of Computer and Information Sciences, 11 (5)

Quinlan J. R., 1983. Learning efficient classification procedures and their application to chess and games. In Michalski J R (Eds.). Machine Learning: An Artificial Intelligence Approach,. Vol. 1, Morgan Kaufmann, San Mateo

Quinlan J. R., 1986. Introduction of decision trees. Machine Learning, 1 (1): 81~106

Quinlan J. R., 1983. C4.5: programs for machine learning. Morgan Kaufmann, San Mateo, CA

Srikant R. and Agrawal R., 1995. Mining generalized association rules. In Proc. of 21st VLDB Conf., Zurich, Switzerland

Srikant R. and Agrawal R., 1996. Mining quantitative association rules in large relational tables. In Proc. of the ACM SIGMOD Conf. on Management of Data, Montreal, Canada

Yager R. R., 1991. On Linguistic summaries of data. In Shapiro G. P. and Frawley W., Eds. Knowledge Discovery in Databases. AAAI Press

Zadeh L. A., 1965. Fuzzy sets. Information And Control, 8: 338~353

Zhang T., Ramakrishnan R. and Livny M., 1996. BIRCH: an efficient data clustering method for very large databases. SIGMOD, 25: 103~114

Zhang T., Ramakrishnan R. and Livny M., 1997. BIRCH: A new data clustering algorithm and its applications. Data Mining and Knowledge Discovery, 1: 141~182

Ziarko W. (ed.), 1994. Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer-Verlag

附录：GIS 术语中英文及其缩写对照表*

中文	英文	缩写
(俄罗斯) 全球轨道导航卫星系统	Global Orbiting Navigation Satellite System	GONSS
(美国) 国际地球科学信息网络中心		CIESIN
(美国) 国家地理信息与分析中心	National Center for Geographic Information and Analysis	NCGIA
(美国) 国家海洋大气局	National Oceanic and Atmospheric Administration	NOAA
(美国) 国家航空与航天局	National Aeronautics and Space Administration	NASA
(美国) 联邦地理数据委员会	Federal Geographical Data Committee	FGDC
(美国) 数字栅格图	digital aster graphics	DRG
(美国) 数字正射影像	digital orthophoto quadrangle	DOQ
(伪) 假-3D GIS	pseudo-3D GIS	
(线) 段	segment	
(中国) 测绘遥感信息工程国家重点实验室	State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing	
(中国) 国家基础地理信息中心	National Geomatics Center of China	NGCC
(中国) 全国地理信息标准化技术委员会	National Technical Committee for Geographical Information	NTCGI
(中国) 资源与环境信息系统国家重点实验室	National Laboratory of Resource and Environment Information System	NLREIS
《地理与地理信息科学》	Geography And Geo-information Science	GAGIS
1984 世界大地坐标系	World Geodetic System for 1984	
2D 字符串	2D-string	
3D 地学信息系统	3D geosciences information system	3D GIS
3D 城市模型	3D city model	3D-CM
3D 地学模拟系统	3D geosciences modeling system	3D GMS
3D 形式化数据结构	3D formal data structure	3D FDS
3D 元件构模	3D component modeling	
4-交模型 (4 元组模型)	4-intersection model	4-I model
9 交模型 (9 元组模型)	9-intersection model	9-I model
Delaunay 三角网 (剖分)	delaunay triangular irregular network	D-TIN
Frame (有限) 差分法	frame finite difference	
ID 副码	auxiliary ID	A-ID
Java 开放式数据库互联	Java Data Base Connectivity	JDBC
JPEG 格式	Joint Photographic Experts Group Format	JPEG

* 本对照表选自本书正文，并参照《地理信息系统名词》(科学出版社，2002)，且摘选了其中的部分重点条目

中 文	英 文	缩 写
Morton 码	Morton code	MT (code)
n-单纯复形	n-complex	
n-维流形	n-pseudomanifold	
p-中心定位问题	p-median location problem	
TIFF 格式	Taggged Image File Format	TIFF
TP 网络模型	tri-prism network model	TPN
凹多边形	re-entrant polygon	
八叉树模型	octree model	
邦联式数据库	federated database	
包含	contain/inclusion	
包含于	in	
贝叶斯最大熵	Bayesian maximum entropy	BME
被穿越	passby	
被覆盖	coverby	
边	edge	
边界/境界	boundary	
边界表示	boundary representation	B-Rep
边界代数多边形充填算法	boundary algebra filling	BAF
边缘/界线	border	
编码规则	encoding rule	
编码模式	encoding schema	
变异函数	variogram	
遍历法	traversal method	
标识符	identifier	
标准通用建模语言	standard general makeup language	SGML
表面	surface	
波特率	band rate	
不闭合多边形	leaking polygon	
不规则三角网	triangular irregular network	TIN
不确定性	uncertainty	
不确定性推理	reasoning under uncertainty	
采样	sampling	
参数估计	parameter estimation	
侧向相邻	lateral spatial adjacency	
层次分析法	hierarchical analysis procedure	AHP
层次数据结构	hierarchical data structure	
层次数据模型	hierarchical data model	
产生	birth	
常规四叉树	common quadtree	CQT
超关系模型	hyper-relational model	
超类	superclass/supertype	
超媒体	hypermedia	

中文	英文	缩写
超三角形	super triangular	
超熵	hyper entropy	H-En
超图数据结构	hypergraph based data structure	HBDS
超文本	hypertext	
超文本建模语言	hypertext makeup language	HTML
朝向侧	face side	
尺度世界	dimensional world	
穿越	pass	
传播	propagation	
传输	transmission	
垂直	vertical	
从数据库中发现知识	knowledge discovery in database	KDD
粗集法	rough sets approach	
粗集法与云理论的结合	rough sets approach plus cloud theory	
大地参照系	geodetic reference system	
大地测量	geodesy	
大地坐标	geodetic coordinates	
单位汇水面积	specific contribution area	SCA
单源点间的最短路径	single source shortest path	
等积地图投影	homolosine map projection	
等价关系	equivalence relation	
等深线	isobath	
等值区域	choropleth	
等值线	isarithmic line	
等值线/等高线/等深线图	contour (chart)	CLM
等值线图	isarithmic map	
地方坐标系	local coordinate system	
地籍	cadastre	
地籍图	cadastral map	
地块	plot	
地理	Geography	
地理（坐标）参照	georeference	
地理编码	geocoding	
地理方位角	geographical azimuth	
地理分布	geo-distribution	
地理空间	geographical space	
地理空间的认知模型	cognitive models of geographic space	
地理空间世界	geospatial world	
地理空间信息学/地球空间信息学	Geomatics	
地理空间元数据库	geospatial meta-database	GMDB
地理实体编码模型	geographical entity encoding model	GEEM
地理相关模型	geo-relational model	

中　文	英　文	缩　写
地理信息科学	Geographical Information Science	GIS
地理信息认知	cognition of geographic information	
地理信息系统	geographic information system	GIS
地理要素集合	feature collections	
地理属性和图形属性耦合的信息编码模型	coupled geographical-cartographic attribute encoding model	CGCAEM
地理坐标	geographical coordinates	
地名学	toponymy	
地球空间信息学	Geo-informatics	
地球球体	earth spheroid	
地球同步轨道	geostationary orbit	
地球同步卫星	geo-synchronous satellite	
地球椭球体	earth ellipsoid	
地球物理	Geophysics	
地球资源观测系统	earth Resource Observation System	EROS
地势图	relief map	
地图比例尺	map scale	
地图代数	map algebra	
地图的分幅号	map sheet numbering	
地图复制	map duplicated	
地图规范	map specification	
地图接边	map adjustment	
地图容量	load of map content	
地图投影变形	map projection distortion	
地图整饰	map appearance	
地图综合	map generalization	
地物类型	feature type	
地形	landform	
地形简化	decimation	
地形数据库	topographic database	
地形图	topographic map	
地形细化	terrain refinement	
地形因子	terrain factor	
地学	Geosciences	
地学数据	earth science data, geo-data	
地学信息	geo-information	
地学信息处理	geoprocessing	
地学信息系统	geosciences information system	GIS
地质	Geology	
地质空间	geological space	
地质统计学	Geostatistics	
地质细胞/元胞	geocellular	

中　文	英　文	缩　写
第一范式	1NF	
点/地理点列	point	
点实体	point entity	
电子地图	electronic map	
叠置/叠加	overlay	
定位精度	positional accuracy	
定位误差	position error	
定义集	definable set	
段	section	
断面-三角网混合	section-TIN mixed	
断面线	section lines	SLs
断面序列/序列断面	series sections	
对地观测系统	earth observation system	
对地静止卫星	geostationary satellite	
对象/目标	object	
多边形检索	polygon retrieval	
多边形内点判断	point-in-polygon operation	
多边形网格模型	polygon mesh model	PMM
多媒体	multimedia	
多面体	polyhedral/polyhedron	
多态性	polymorphism	
二阶（有限）差分法	2nd finite difference	
二进制	binary	
二维行程	two dimensional run-length	2DRL
方差	variance	
方法学	Paradigm	
方位/坡向	aspect	
方向矩阵	direction relation matrix	DRM
非第一范式		N1NF
非规则块体模型	irregular block model	
非监督学习	unsupervised learning	
分布式数据库	distributed database	
分割	fragmentation	
分类	classification	
分类体系	taxonomy	
分裂	split	
分裂容限	split-tolerance	
分配法	collation method	
封装	encapsulation	
冯罗诺图	Voronoi diagram	
符号化	symbolization	
父节点	parent node	

中　文	英　文	缩　写
覆盖	cover	
概括/归纳/综合	generalization	
概念世界	conceptual world	
高斯-克吕格投影	Gauss-KrÜger projection	
高斯-克吕格坐标	Gauss-KrÜger coordinate	
高斯平面坐标	Gauss plane coordinate	
高斯坐标	Gauss coordinate	
格网	grid	
格网间距	grid interval	
根节点	root node	
更新	updating	
工作区	workspace	
共同边界	conjoint boundary	
关联	conjunction	
关联规则法	association rule approach	
关系结构	relational structure	
广义三棱柱	generalized tri-prism	GTP
归纳算子	induction operator	
归纳学习法	inductive approach	
规范	specification	
规则	rule	
规则格网	regular grids	RGs
规则块体模型	regular block model	
滚筒绘图仪	drum plotter	
国际标准化协会	International Standard Association	ISA
国际标准化组织	International Standard Organization	ISO
国际标准化组织地理信息标准化委员会	Technical Commission 211 of ISO	ISO-TC211
国际地理学联合会	International Geographical Union	IGU
国际电工委员会	International Electricity Commission	IEC
国际摄影测量与遥感学会	International Society for Photogrammetry and Remote Sensing	ISPRS
国际制图协会	International Cartographic Association	ICA
国家地理空间数据交换中心	National Geographical Data Center	NGDC
国家空间数据基础设施	national information infrastructure	NII
过伸	overshoot	
含象限前缀的主副码耦合编码方法	quadrant label prefixed for principal code and auxiliary code	QUAPA
合并/崩溃	collapse	
合成集	composed set	
横轴墨卡托投影	inverse Mercator map projection	
后	after	

中 文	英 文	缩 写
弧	arc	
互操作	interoperability	
缓冲区	buffer	
绘图板	drawing board	
混合模型	mixed model	
霍夫曼变换	huffuman transformation	
基本集	elementary set	
基态修正模型	base state with amendments model	
基于 Voronoi 图的 9I 模型	Voronoi 9-intersection model	V9I model
基于对象的	object-based	
基于图像的绘制	image-based rendering	
基于图像的建模	image-based modeling	
基于网络的	network-based	
基于位置的服务	location-based services	LBS
基于现象的	phenomenon based	
基于域的	field-based	
基于状态和变化的统一时空数据模型	State-and-change-based unified data model	SCUDM
基元/单形/图元	primitive	
集/集合	set	
集成	integration	
几何特征	geometry	
几何元素	geometric elements	
继承	inheritance	
加拿大标准协会	Canadian Standards Association	CSA
加拿大地理信息系统	Canada Geographical Information System	CGIS
加拿大遥感中心	Canada Center for Remote Sensing	CCRS
间曲线	intermediate contour line	
间跳	popping	
监督学习	supervised learning	
剪切	clip	
剪切窗口	clipping window	
简单差分法	simple difference	
简化	simplification	
交叉方法	intersection-based model	
交互方法	interaction-based method	
交互式编辑	interactive editing	
交互式算法	interchange algorithms	
交换格式	interchange format	
节点	node	
节点吻合	node snap	
结构化查询语言	Structured Query Language	

中文	英文	缩写
结构实体几何	construction solid geometry	CSG
金字塔技术	mip-map technology	
金字塔模型	pyramid model	
近似空间	approximation space	
进度模型	progressive mesh	PM
进化	evolution	
进入	enter	
经度	longitude	
景观	landscape	
局部优化方法	local optimization procedure	LOP
聚分性	aggregation/subdivision	
聚集/聚合	aggregation	
聚类	clustering	
聚类法	clustering approach	
聚类分区	cluster zoning	
聚类分析	cluster analysis	
决策树	decision tree	
均方根点位位置误差	root mean square position error	
开发式地理信息系统	Open GIS	
开放地理信息系统协议	OpenGIS Consortium	OGC
开放系统互联参考模型		OSI/RM
开放性地理数据互操作规范	Open Geodata Interoperation Specification	OpenGIS
开放性数据库互连	Open Data Base Consortium	ODBC
科学可视化	scientific visualization	
可持续发展	sustainable development	
可视化	visualization	
可视网络	visibility network	
可视性分析	visibility analysis	
可视域	viewshed	
可选的	optional	O
客户/服务器	client/server	C/S
空间参照系	spatial reference system	
空间分布规律	spatial distribution regularities	
空间分类规则	spatial classification regularities	
空间关联规则	spatial association regularities	
空间几何事件	spatial geometry event	SGE
空间邻近	spatial adjacency	
空间区分规则	spatial discriminate rules	
空间数据仓库	spatial data warehouse	
空间数据库级空间元数据库	general spatial metadata database	GSMD
空间数据挖掘	spatial data mining	SDM
空间数据挖掘与知识发现	spatial data mining and knowledge discovery	SDMKD

中　文	英　文	缩　写
空间数据引擎	spatial data engine	SDE
空间数据质量	spatial data quality	
空间特征规则	spatial characteristics rules	
空间拓扑事件	spatial topological event	STE
空间为主	space-primary	
空间相邻	spatial adjacent	
空间演化规则	spatial evolution regularities	
快照	snapshot	
矿山 GIS	mine GIS	MGIS
矿山数据挖掘	mine data mining	MDM
兰勃特等积方位投影	Lambert's equal-area meridional map projection	
类	class	
类别	category	
类三棱柱模型	analogical tri-prism model	ATP model
棱柱体模型	prismatic model	
历史记录	historical record	
连接结点	connected node	
连通性	connectivity	
联合/关联	association	
链	link	
邻接	connection	
邻接分析	adjacency analysis	
邻接区域	adjacent area	
邻接图幅	adjacent sheets	
邻近分析	proximity analysis	
邻域	neighbor/neighborhood	
邻域分析	neighborhood analysis	
邻域路径	neighborhood paths	
邻域算子	domain operator	
邻域图	neighborhood graphs	
流式数字化	stream node digitizing	
流向驱动法	aspect-driven	
流域网络	drainage network	
六面体	hexahedral/hexahedron	
陆地流水路径	land water routine	
路径	route	
路径分析	route analysis	
路径分析算法	routing algorithm	
路径系统	route-system	
略图	outline map	
马氏距离	Mahalonobis distance	

中 文	英 文	缩 写
美国地理信息科学大学研究会	University Consortium for Geographic Information Science	UCGIS
美国地质调查局	United States Geological and Survey	USGS
面	face	
面/面域/区域	area	
面模型	facial model	
面实体	surface entity	
面条数据模型	spaghetti data model	
面向对象的	object-oriented	OO
面向对象的 3D 模型	object oriented 3D model	OO-3DM
面向对象的知识	object oriented knowledge	
面向对象数据模型	object-oriented data model	
面向属性的诱导学习法	attribute oriented induction	AOI
命名法	denomination method	
模糊分析	fuzzy analysis	
模糊集理论	fuzzy sets theory	
模块	module	
模块化软件	modular software	
模式识别	pattern recognition	
目标	target	
目标区	target region/target area	
目录	catalogue	
目视判读	visual interpretation	
内部	interior	
内部参考框架	intrinsic reference frame	
内插区	zone of interpolation	
内多边形	hole	
欧几里德距离	Euclidean distance	
欧洲标准化委员会	Comite Europeen de Normalisation (法)	CEN
欧洲标准化委员会地理信息技术委员会	Technical Committee For Geographical Information, CEN	CEN/TC287
欧洲地图事务组织		MEGRIN
欧洲空间局	European Space Agency	ESA
配准	registration	
批	batch	
匹配	matching	
偏差	deviation	
平滑	smoothing	
平面点集 Voronoi 图	ordinary Voronoi diagram	OVD
平面图	planimetric map	
平面坐标	planar coordination	
坡向分析	aspect analysis	

中　文	英　文	缩　写
坡向图	aspect map	
破碎多边形/无意义多边形	sliver polygon	
普遍的几何知识	general geometric knowledge	
期望值	expected value	Ex
启发法	drop heuristic	
启发式算法	heuristic algorithms	
起点	start point	
区域级空间元数据库	district spatial metadata database	
曲线拟合	curve fitting	
趋势分析	trend analysis	
全国地理信息标准化技术委员会（中国）		CSBTS/TS230
全局级空间元数据库	global spatial metadata database	
全球导航卫星系统	global Navigation Satellite System	GNSS
全球定位系统	Global Positioning System	GPS
全球空间数据基础设施	Global Spatial Data Infrastructure	GSDI
全球数字地图	digital Chart of the world	DCW
全球综合观测系统	integrated Global Observation System	IGOS
全数字化测图	fully digital mapping	
确定性	validity	
确定性 8 点法	deterministic eight-node	D8
认知	cognition	
冗余	redundancy	
三角网数字地形模型	triangulated digital terrain work	
三阶（有限）差分法	3rd finite difference	
三棱柱模型	tri-prism model	TP
三维地理矩阵	3D geographic matrix	
扫描仪	scanner	
熵	entropy	En
上近似	upper approximation	
摄影测量	Photogrammetry	
神经网络法	neural network approach	
神经元网络	artificial neural network	
十六进制的	hexadecimal	
时空复合模型	spatial-time composite model	
时空数据	spatio-temporal data	
时空数据库	spatial-temporal database	STDB
时态 GIS	temporal GIS	TGIS
时态数据库	temporal database	TDB
时态特征	temporal characteristic	
识别框架	frame of discriminate	
实时	real time	

中　文	英　文	缩　写
实体	entity	
实体/体	solid	
实体关系模型	entity relationship model	E-R model
实体关系图	entity relationship diagram	ERD
实体实例	entity instance	
矢量	vector	
矢量化	vectorization	
矢量-栅格转换	vector-to-raster conversion	
世界大地坐标系	world geodetic system	
事件	event	
事件时间	event time	
视点	viewpoint	
视见区	viewport	
视相关 PM 法	vision dependence PM	VDPM
数据仓库	data warehouse	
数据操作	data manipulation	
数据共享	data sharing	
数据交换网络/数据中心	clearinghouse	
数据考古	data archaeology	
数据库链接	database link	
数据链系统	data string system	DSS
数据模型	data model	
数据挖掘	data mining	
数据挖掘与知识发现	data mining and knowledge discovery	DMKD
数据现势性	data currency	
数据字典	data dictionary	DD
数学期望曲线	mathematical expected curve	MEC
数学形态学	Mathematic Morphology	
数字表面模型	digital surface model	DSM
数字地貌模型	digital geomorphic model	DGM
数字地球	digital Earth	
数字地形模型	digital terrain model	DTM
数字高程模型	digital elevation model	DEM
数字化板	digitizing tablet	
数字化地图	digitized map	
数字化仪	digitizer	
数字化仪坐标	tablet coordinates	
数字矿山	digital mine	DM
数字摄影测量	digital photogrammetry	
数字图像处理	digital image process	DIP
数字正射影像	digital orthoimage/orthoimagery/orthophoto	
数字正射影像图	digital orthophoto map	DOM

中　文	英　文	缩　写
双边界直接搜索法	double boundary direct finding	DBDF
双线性内插	bilinear interpolation	
双向反射系数	bia-direction reflection factor	BRDF
水系图	drainage map	
四半无限区域	four semi infinite area	FSIA
四叉树	quatree	Q-tree
四面体	tetrahedral/tetrahedron	
四面体格网	tetrahedral network	TEN
四元三角网	quaternary triangular mesh	QTM
随机八方向（算法）	random eight-node	Rho8
随机四方向（算法）	random four-node	Rho4
所有点对之间的最短路径	all pairs shortest path	
泰森多边形	Thiessen polygon	
贪婪插入法	greedy insertion	
探测	snooping	
探测分析法	exploratory data analysis	EDA
探测性空间分析技术	exploratory spatial analysis	ESA
探测性诱导学习方法	exploratory inductive learning	EIL
特征规则	characteristic rules	
特征为主	feature primary	
提取	extraction	
体模型	volumetric model	
体实体	body entity	
体素	voxel	
条件必须的	conditional	C
通视性	intervisibility	
通用横轴墨卡托投影	Universal Transverse Mercator	UTM
通用数据结构	versatile data structure	
统计法	statistical approach	
统计分析	statistical analysis	
凸包/凸壳	convex hull	
凸多边形	convex polygon	
凸集	convex	
图斑模型	piecewise model	
图斑综合	patch generalization	
图边	margin	
图表数据	tabular data	
图层	coverage	
图幅	mapsheet	
图幅范围	map extent	
图幅拼接	map join	
图廓	map border	

中　文	英　文	缩　写
图例	Legend	
图像处理	image processing	
图像分析与模式识别	image analysis and pattern recognition	IAAPR
图像判读	image interpretation	
图像配准	image registration	
图像融合	image fusion	
图形交换格式	Graphic Interchange Format	GIF
图形实体编码模型	cartographic entity encoding model	CEEM
土地覆盖	land cover	
拓扑（学）	topology	
拓扑关联数据库	topological linked database	
拓扑关系	topological relationship	
拓扑结构化数据	topologically structured data	
拓扑矢量边界表示	topological vector boundary representation	TVBR
拓扑数据	topological data	
外部	exterior	
外部参考框架	extrinsic reference frame	ERF
外推	extrapolation	
完整性	integrity	
网格单元/像元/像素	cell	
网络结构	network structure	
网络数据模型	network data model	
维数扩展	dimension extension	DE
伪多边形	pseudo polygon	
伪节点	pseudo node	
纬度	latitude	
未及	undershoot	
位图	bit map	
位置误差	location error	
纹理	texture	
纹理元素	texel	
系统时间	system time	
细胞自动机/单元自动演化（算法）	cellular automation	CA
细节层次（模型）	level of detail	LOD
下近似	lower approximation	
现实世界	real world	
线	line	
线段交叉	line intersection	
线划图	line graph	
线框模型	wire frame model	
线实体	line entity	
线性四叉树	linear quadtree	LQT

中　文	英　文	缩　写
线与多边形叠置	line overlay	
限定四叉树	restricted quadtree	
相等/同时	equal	
相关性	relativity	
相交	cross	
相接	touch	
相离	disjoint	
相遇	meet	
镶嵌	tessellation	
镶嵌式数据模型	tessellation data model	
项目世界	project world	
像元/像素	pixel	
像元分辨率	cell resolution	
消亡	death	
消息	message	
协议	protocol	
信息	information	
信息发现	information discovery	
信息收获	information harvesting	
形态学	Geomophics	
修正/更改	modification	
虚拟现实	virtual reality	
虚拟云	virtual cloud	
悬挂节点	dangle node	
压缩	compression	
亚洲遥感研究信息网	Asia Remote Sensing Research Information Network	ARSRIN
严重缺陷	serious defect	serious - def
遥感	Remote Sensing	RS
要素/特征	feature	
要素类/地物类	feature class	
叶节点	leaf node	
一般缺陷	light - defect	light-def
一体化数据结构	unified data structure	
一维行程	one dimensional run-length	1DRL
一致性	consistency/conformance	
遗传基因法	genetic algorithms	GA
遗漏性误差	omission error	
异常值	outlier	
隐含线消除	hidden line removal	
永久对象	permanent object	
用户标识码	user-ID	
用户需求分析	user requirement analysis	

中 文	英 文	缩 写
游程（长度）编码	run-length coding	
有向图	digraph	
有序的	ordinal	
诱导学习法	inductive learning	
语义信息	semantic information	
预测	prediction	
域/字段	field	
域/字段	region	
阈值	threshold	
元变量	metavariable	
元程序	metaprogram	
元代码	metacode	
元调用	metacall	
元符号	metasymbole	
元概念	metanotion	
元规则	metarule	
元数据	metadata	
元数据	metadata	
元数据元素	metadata element	
元数据字典	metadata dictionary	MDD
元系统	metasystem	
元字符	metaword	
原型	prototype	
圆标准误差	circular standard error	
圆概率误差	circle probable error	
圆柱投影	cylindrical projection	
源区	source region	
约束	constraint	
约束 Delaunay 三角网（剖分）	constrained delaunay triangular irregular network	CD-TIN
云变换	cloud transform	
云发生器	cloud generator	
云理论	cloud theory	
云模型	cloud model	
增强	enhancement	
增强模式	enhanced mode	
栅格	raster	
栅格-矢量转换	raster-to-vector conversion	
折点/顶点	vertex	
针体模型	needle model	
真 3D GIS	real-3D GIS	
证据理论	evidence theory	
政府协议体系	government OSI profiles	GOSIP

中　文	英　文	缩　写
知识	knowledge	
知识提取	knowledge extraction	
直方图	histogram	
直接参考框架	direct reference frame	
值域	value domain	
植被指数	vegetation index	
质心/形心/中心	centriod	
中国 GIS 协会	China Association for GIS	CAGIS
中国测绘科学院	China Academy of Surveying and Mapping	CASM
中国测绘学会	China Society for Geodesy, Photogrammetry and Cartography	CSGPC
中国地理空间数据协调委员会	China Interagency Coordinating Committee on Geo-spatial Data	CICCGD
中国地理学会	Geographical Society of China	
中国国家遥感中心	National Remote Sensing Center of China	NRSCC
中国海外地理信息系统协会	Association of Chinese Professionals in GIS (Abroad)	CPGIS
中国科学院地理科学与资源研究所	Institute of Geographic Sciences and Natural Resources Research, CAS	IGSNRR
中国科学院遥感应用研究所	Institute of Remote Sensing Application, CAS	IRSA
中国全球定位系统应用协会	China Technical Application Association for GPS	CTAAGPS
中国图像图形学会	China Society for Image and Graphics	CSIG
中国遥感学会	China Society for Remote Sensing	CSRS
中国资源卫星应用中心	China Center for Resource Satellite Data and Application	CCRSDA
中心点	center point	
中心透视	center perspective	
中心线	center line	
中央子午线	center meridian	
终点	end point	
终节点	terminating node	
重叠	overlap	
重缺陷	heavy defect	heavy-def
主方向	ordinal directions	
主码	principal ID	P-ID
属性	attribute	
属性事件	attribute event	AE
专家系统	experts system	ES
专题	thematic	
专题图	thematic map	
专用标准/剖面	profile	
状态	state	

中文	英文	缩写
锥	cone	
准-3D GIS	quasi-3D GIS	
子类	subclass/subtype	
宗地	cadastral parcel	
宗地图	cadastral plan	
宗地图	parcel map	
总汇水面积	total contribution area	TCA
族	family	
阻强	impedance	
组合算子	combination operator	
最大坡降法	maximum drop slope	
最短路径跟踪算法	minimum path tracing algorithm	
最邻近拓扑邻接图	closest topological relationship graph	CTRG
最小边界长方体化表示	minimum boundary cubic representation	MBCR
最小外接矩形	minimum enclosing rectangle	MER
最小外接四边形	enclosing rectangle	
最小约束矩形	minimum bounding rectangle	MBR
坐标变换	coordinate conversion	
坐标转换	coordinate transformation	

后　记

时间过得真快，从2002年6月开始写作，到今天已经一年零两个月。期间，我们熬过了香港盛夏的潮热，经历了北京严冬的寒冷，还有那个令京港两地人心惊胆战的春夏。终于，完成了本书的初稿、修改、审阅、校对等全部工作，该是坐下来回味一下本书从酝酿到诞生全过程的经历的时候了。

早在1996年开始给研究生讲授“GIS原理”课程时，即感觉到适合我国研究生使用的GIS教材的缺乏，于是萌生了积累素材编写一部关于“GIS原理与算法”书的愿望。到1999年，素材积累基本完成，多媒体教学课件也经过了4年使用和反复修改，取得了满意的教学效果，研究生们纷纷敦促尽早出版本教材。由于前几年科研、教学任务过于紧张，一直没有整段的时间坐下来了此心愿。此事一拖就是3年。

2002年夏天，史文中教授邀请我到香港理工大学合作4个月（6～10月），真是良机难得，终于可以排除各类纷扰，静静地坐下来“闭关修炼”了。这个夏秋季节，我与史文中教授携手，从本书的大纲拟定、章节安排、内容筛选等各项细节进行了认真的讨论和仔细推敲，对国内外GIS原理与算法方面的最新研究成果进行了系统的梳理总结和归纳演绎。兼顾内容的基础性、前沿性、成熟性和系统性原则，充分结合京港两地研究生GIS课程教学经验与心得体会，并融入我们两个小组近年在本领域的研究成果，完成了本书的初稿。

在初稿写作过程中，GIS界的许多朋友为我们提供了无私的帮助（如提供文献资料），并进行了多次有益的讨论。他们是：香港中文大学的林珲教授，香港理工大学的李志林教授，武汉大学的龚健雅、李清泉、王新洲和陈晓琳教授，解放军信息工程学院的朱长青和朱述龙教授，同济大学的刘大杰教授、刘春博士等。本书初稿完成后，中国矿业大学的郭达志教授对初稿进行了全面审阅，并提出了一些建设性建议；中国矿业大学2003年“GIS原理与算法”课程的40余名博士生、硕士生在书稿试用过程中分别发现、指出和纠正了其中的多处错误；博士生张文胜、陈学习、王彦兵先后对论文初稿与修改稿进行了阅读和校核，硕士生陈书琳绘制了书中的部分图表，硕士生贾晓琳对书中GIS术语进行了中、英文及缩写对照表整理。在此，一并表示感谢！此外，本书的顺利出版，还和许多未曾提及的亲人和朋友的关心与支持分不开。对于所有为本书付出过智慧、劳动和关爱的人们，道一声：谢谢！

应该说，本书是国际GIS理论与算法研究的集大成，是集体成果的结晶。写作过程中，参考了国内外大量的书籍、刊物、会议论文和学位论文，其成果有的在书中直接引用并做了注明，有的在书中引用了而由于疏漏却未作注明，凡此均请见谅！

本书就要出版了，却突感心中莫名，如同母亲在孩子即将降生时的那种复杂心情，既有对希望的期盼又有对缺憾的担忧。期盼的是所付出的辛劳能使绝大多数读者满意，担忧的是限于篇幅而遗漏了许多高水平研究成果。GIS理论、技术与应用的不断发展，本身就是挑战，就是激励。希望我们能按照李德仁院士的嘱咐，“继续关注本领域国内

外研究动态和最新成果，随时对新理论、新算法进行总结、归纳和提炼，以便在本书后续再版中不断纳入进去，使本书的体系和内容不断丰富和完善”。同时，恳切地期待GIS学界的方家、同仁多批评、指教，为本书的补充与完善提出建议，使得本书能与时俱进。

作者

于北京、香港

2003年8月2日

［Ｇｅｎｅｒａｌ　Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝地理信息系统原理与算法
作者＝
页数＝４４１
ＳＳ号＝０
出版日期＝